What Is Judicial Ideology, and How Should We Measure It?

Joshua B. Fischman*
David S. Law**

INTRODUCTION

It has become de rigueur for leading law schools to profess great enthusiasm for both interdisciplinary and empirical research. Yet not all work in this vein has been warmly embraced. There remains deep skepticism in legal circles toward interdisciplinary empirical scholarship aimed at capturing the impact of ideology on judicial behavior. Judge Harry Edwards of the D.C. Circuit, a vocal critic of this body of work, has vigorously disputed that "'ideology' broadly influences decision making."[1] The "disciples" of what he calls the "political view," he writes, "seem determined to characterize judges as knee-jerk ideologues, who act pursuant to a blind adherence to ideological precepts and decide cases wholly without regard to the

* Associate Professor of Law, University of Virginia. A.B. (1994), Princeton University; J.D. (1999), Yale Law School; Ph.D. (Economics) (2006), M.I.T.

** Professor of Law and Professor of Political Science, Washington University in St. Louis. B.A. (1993), M.A. (2000), Ph.D. (Political Science) (2004), Stanford University; J.D. (1996), Harvard Law School; B.C.L. in European and Comparative Law (2003), University of Oxford. Earlier versions of this Article were presented at the Symposium on Empirical Research on Decision-Making in the Federal Courts held at Washington University in St. Louis in April 2008, the Conference on Empirical Legal Studies held at Cornell Law School in September 2008, and the Law, Economics, and Politics Colloquium at NYU Law School. We thank the participants on all three occasions for their insightful comments and suggestions. For their detailed feedback, we are particularly grateful to Lewis Kornhauser, Greg Mitchell, Greg Sisk, and Brian Tamanaha. Our special thanks to Andrew Martin and the Center for Empirical Research in the Law for organizing the symposium at Washington University, and to the editors of the *Washington University Journal of Law & Policy* for their patience and assistance in bringing this Article to fruition.

[1] Harry T. Edwards, *Collegiality and Decision Making on the D.C. Circuit*, 84 VA. L. REV. 1335, 1336 (1998).

law."[2] "Political scientists who study the Supreme Court do not take legal doctrine very seriously," charges Michael Dorf, a prominent constitutional scholar.[3] In suggesting that ideology influences the behavior of the Justices, he argues, political scientists have been guilty of "dispens[ing] with the metaphysical nonsense of law as a category independent of values, ideology and preferences, at least in the sorts of hard cases that reach the Supreme Court."[4] Brian Tamanaha, a leading legal theorist, is no more generous in his assessment: "The judicial politics field," he charges, "was born in a congeries of false beliefs that have warped its orientation and development," and it remains characterized by "a distorting slant" that leads scholars "to exaggerate the influence of politics in judging."[5]

Has empirical research on the influence of judicial ideology gone awry? If so, how? To be sure, the fault may not lie entirely with the literature itself. To some extent, there exists a straightforward problem of "unfortunate," and innocent, "interdisciplinary ignorance."[6] It is understandably difficult for legal scholars and judges to navigate and evaluate a body of literature that, although relevant, tends to be disseminated through unfamiliar channels and can often be highly technical.[7] There is also little reason to expect those who practice or teach the craft of legal argument to embrace a body of research that questions the extent to which judicial decision-making is actually driven by legal argument.[8] Empirical work that portrays ideology as an important determinant of judicial behavior

[2] Harry T. Edwards, *Public Misperceptions Concerning the "Politics" of Judging: Dispelling Some Myths About the D.C. Circuit*, 56 U. COLO. L. REV. 619, 625 (1985).

[3] Michael C. Dorf, *Whose Ox Is Being Gored? When Attitudinalism Meets Federalism*, 21 ST. JOHN'S J. LEGAL COMMENT. 497, 498 (2007).

[4] *Id.* at 499.

[5] Brian Z. Tamanaha, *The Distorting Slant of Quantitative Studies of Judging* 4 (St. John's Univ. Legal Studies Research Paper Series, Paper No. 08-0159, 2008), *available at* http://ssrn.com/abstract=1292459; *see also* BRIAN Z. TAMANAHA, BEYOND THE FORMALIST-REALIST DIVIDE ON JUDGING: HISTORY, POLITICAL SCIENCE, THEORY (forthcoming 2009).

[6] *See generally* Frank B. Cross, *Political Science and the New Legal Realism: A Case of Unfortunate Interdisciplinary Ignorance*, 92 NW. U. L. REV. 251 (1997).

[7] *Cf. id.* at 254 (observing that "legal scholars may ignore political science research because it is inconvenient").

[8] *See id.* (noting that lawyers have a "considerable financial stake in perpetuating" the belief that legal doctrine explains how judges decide cases).

breaches the wall of separation between law and politics that legal scholars have labored mightily for decades to erect and defend.[9] If legal and political decision-making come to be seen as largely undifferentiated, it becomes unclear why judges should pay any special heed to legal as opposed to policy arguments; nor, for that matter, does it remain obvious why certain questions should be resolved in the courts rather than in the arena of ordinary politics. Breaches in the wall between law and politics therefore threaten to diminish both the range of policy questions over which legal scholars may attempt to claim special expertise and the extent of the influence that they have over the determination of those questions.

But skepticism of the empirical literature on judicial ideology cannot be wholly dismissed as the product of ignorance or self-interest on the part of the audience. Without seeking to deny that ideology plays a significant role in judicial decision-making, well-informed observers have nevertheless raised reasonable criticisms about the manner in which empirical scholars have tackled the subject.[10] Critics of the empirical social science literature on judicial behavior have reason for concern, in particular, about the manner in which researchers have sought to measure judicial ideology. Empirical studies routinely purport to measure ideology without specifying what is meant by "ideology," or taking care to measure "ideology" in a way that will not invite a host of objections. Given the skepticism of legal scholars toward this line of research, however, it is incumbent upon those of us who investigate judicial ideology to employ concepts and methods that are both clear and appropriate. The fact that much of the audience is not methodologically

[9] *See, e.g.*, DUNCAN KENNEDY, A CRITIQUE OF ADJUDICATION [FIN DE SIÈCLE] 14, 23–38 (1997) (arguing that legal and political discourse deny the extent to which the ideological tendencies of judges resolve legal questions, and critiquing the traditional view that adjudication, unlike legislation, "need not be political").

[10] *See, e.g.*, Barry Friedman, *Taking Law Seriously*, 4 PERSPECTIVES ON POL. 261, 262 (2006) (admonishing political scientists for exhibiting "an almost pathological skepticism that law matters" and for failing "to take law and legal institutions seriously"); Patricia M. Wald, *A Response to Tiller and Cross*, 99 COLUM. L. REV. 235, 239–40, 250 (1999) (acknowledging freely that "judges do have personal ideologies which sometimes enter into their decisionmaking," and questioning whether that fact can be "legitimately labeled a problem" at all, yet also deeming it "extremely problematic" to assume, as scholars have sometimes appeared to do, "that judges intentionally act in alignment with the party from which they sprung").

sophisticated makes it all the more crucial that we do so. As a research community, we must cultivate and convey a better understanding of methods for measuring judicial ideology if we are to succeed in convincing others of the validity of our work.

Empirical scholarship on the subject of judicial ideology is vulnerable to two sets of difficulties, which tend to blend into one another. The first set is theoretical; the second set is methodological. Both will be surveyed here. Part I of this Article explores the theoretical problem that scholars use the term "judicial ideology" in the absence of any widespread agreement or clear understanding as to what the term means in the first place. It is difficult for scholars to devise appropriate and broadly acceptable measures of judicial ideology when they and their readers have different concepts—or perhaps no coherent concept at all—of "judicial ideology" in mind. As a result, bona fide intellectual disagreement over the nature of judicial behavior is too easily compounded by outright misunderstanding. Part II discusses three of the most significant and common practical obstacles to the measurement of judicial ideology. First, ideology is not a tangible phenomenon that can be directly observed. Second, judicial behavior is often open to multiple interpretations. Third, judicial ideology may be a multidimensional phenomenon, such that a judge who is liberal in one context may be moderate or conservative in another, or the labels "liberal," "moderate," and "conservative" may not seem applicable at all.

Parts III and IV of this Article aim to address an important practical need of judicial behavior scholars. There are several ways in which one can measure judicial ideology, but little has been written about how researchers should go about choosing among these methods. Competing considerations of accuracy, convenience, and ease of interpretation make it difficult to know what measurement approach is most appropriate to the task at hand. In Part III, we identify and discuss the relative merits of three popular approaches: the use of proxy measures, the assessment of judicial ideology based on the actual behavior of the judges in a particular context, and the transplantation of ideology measures developed in one context into other contexts involving partly or wholly different data.

Finally, in Part IV, we compare the real-world performance of different measurement approaches, using data drawn from the federal courts of appeals and the Supreme Court. Our comparisons establish that efforts to estimate the impact of ideology on judicial behavior can yield significantly different results depending upon how one chooses to measure ideology. Ultimately, there is no panacea for the many challenges involved in identifying a viable and appropriate measure of judicial ideology for the research question at hand. In light of their superior performance in our tests, however, we believe that ideology measures derived from the actual behavior of judges may deserve more serious attention and wider usage than they have thus far received, particularly as compared to the proxy measures that remain popular among scholars.

I. THEORETICAL CHALLENGES

Scholars have used the term "ideology" in a bewildering variety of ways, often without even attempting to define it.[11] The result is that ideology is in the eye of the beholder: what one observer might call ideological behavior, another might call principled judging, and vice versa. An ideology is, in a literal sense, a collection or system of ideas. The *Oxford English Dictionary* offers four definitions of the term, only one of which corresponds to the sense in which legal scholars and social scientists use the term in connection with judicial behavior: "ideology" refers, in this sense, to "[a] systematic scheme of ideas, usu[ally] relating to politics or society, or to the conduct of a class or group, and regarded as justifying actions, esp[ecially] one that is held implicitly or adopted as a whole and maintained regardless of the course of events."[12]

[11] *See, e.g.*, Philip E. Converse, *The Nature of Belief Systems in Mass Publics*, *in* IDEOLOGY AND DISCONTENT 206, 207 (David Apter ed., 1964) (observing, a quarter century ago, that the term "ideology" had already been muddled by diverse uses, "and opting therefore to employ the term "belief system" instead); Bryan D. Lammon, *What We Talk About When We Talk About Ideology: Judicial Politics Scholarship and Naïve Legal Realism*, 83 ST. JOHN'S L. REV. (forthcoming 2009), *available at* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1264174 (noting that "rarely does judicial politics scholarship pause to investigate what it means by ideology"); David W. Minar, *Ideology and Political Behavior*, 5 MIDWEST J. POL. SCI. 317, 320–26 (1961) (discussing and comparing the most prominent ways in which the term "ideology" has been employed in social science).

[12] OXFORD ENGLISH DICTIONARY (2d ed. 1989), http://dictionary.oed.com; *see also, e.g.*, Minar, *supra* note 11, at 320–26 (reviewing various conceptions of ideology that have been commonly used by social scientists).

Defined in this manner, however, "ideology" is not amenable to empirical research.[13] With the tools and data that currently exist, there is simply no way for researchers to observe directly a judge's actual state of mind.[14] As a result, studies that purport to measure "judicial ideology" have not sought to ascertain the structure or content of the ideas that judges hold. In practice, they have aimed instead to measure the extent to which judges behave in a way that appears motivated by preference or beliefs of an ideological character. Yet this orientation of the literature toward observable behavior, as opposed to unobservable states of mind, still begs the question of what it means for a judge to behave "ideologically."

Even in the relatively specific context of empirical research on judicial behavior, there is an array of different phenomena that people may have in mind when using the term. It might be considered "ideological," for example, for judges to seek to advance a particular policy outcome—a world characterized by less environmental degradation, or of less regulation, or of greater or lesser levels of immigration. Alternatively, the term "ideological" could describe a tendency to favor or disfavor certain types of parties—criminal defendants, police officers, corporations, members of ethnic or religious minorities, the disabled, and so forth. Indeed, the breadth of the concept of "ideology" even makes it possible to speak of the existence of both *political* ideology and *legal* ideology. To say that a certain type of judicial behavior is "ideological" need not mean that it is ideological in a *political* sense: one might, for example, characterize adjudication that relies heavily upon logical deduction from formal rules as narrowly "legal," whereas adjudication driven by ideas about the role of law and the responsibilities of judges might by contrast be characterized as both "legal" and "ideological" in character.

The existing literature, however, tends to equate judicial ideology with political orientation, and to distinguish sharply between political

[13] *See* Minar, *supra* note 11, at 320 (noting that efforts to study the relationship between ideology and political behavior are complicated in part by tensions between the way in which the term has traditionally been used and "the requirements of scientific explanation").

[14] *See infra* Part II.A (discussing the methodological challenge that a psychological phenomenon such as the ideology of a judge cannot be directly observed using currently available techniques).

and ideological motivations, on the one hand, and legal motivations, on the other. The idea that "politics" and "ideology" are synonymous, and that "law" and "ideology" are opposites, forms the basis of the theoretical framework articulated by Professors Segal and Spaeth, who assign explanations of judicial decision-making to one of three models: the "legal model," the "attitudinal model," and the "strategic" or "rational choice model."[15] The "attitudinal model," of which they are the chief proponents,[16] equates "the ideological attitudes and values of the justices" with their political leanings.[17] As they use the term, ideology refers simply to the political leanings of the Justices.[18] What they call the "legal model," by contrast, depicts judicial decision-making as the product of the interplay of law and fact.[19]

This distinction between the "legal" and the "political" or "ideological" is deeply problematic. To take a concrete example, Justice Scalia has articulated at length a set of reasons for favoring adherence to the original meaning of the text as a method of statutory and constitutional interpretation.[20] Most legal and political observers alike would conclude that it is appropriate, if not desirable, for him to adopt an interpretive method, or "judicial philosophy," in light of his responsibilities as a judge. Yet that is not to say that his choice of originalism is neither "ideological" nor "political" in character. First, it is possible that he has given an incomplete account of his reasons for choosing a particular judicial philosophy. Might it be that his

[15] *See* JEFFREY A. SEGAL & HAROLD J. SPAETH, THE SUPREME COURT AND THE ATTITUDINAL MODEL REVISITED (2002) (contrasting the "attitudinal" model with the "legal" and "rational choice" models); *see also, e.g.*, RICHARD A. POSNER, HOW JUDGES THINK 19–56 (2008) (identifying nine theories of judicial behavior, including the "attitudinal," "strategic," and "legalist" theories).

[16] *See* FORREST MALTZMAN ET AL., CRAFTING LAW ON THE SUPREME COURT: THE COLLEGIAL GAME 10–13 (2000) (implying that Professors Segal and Spaeth may not be in the mainstream of judicial politics research insofar as they reject the position that strategic considerations influence the voting behavior of Supreme Court Justices).

[17] "[T]he Supreme Court decides disputes in light of the facts of the case vis-à-vis the ideological attitudes and values of the justices. Simply put, Rehnquist votes the way he does because he is extremely conservative; Marshall voted the way he did because he was extremely liberal." SEGAL & SPAETH, *supra* note 15, at 86.

[18] *See id.*

[19] *See id.* at 48.

[20] *See* ANTONIN SCALIA, A MATTER OF INTERPRETATION: FEDERAL COURTS AND THE LAW 9–25, 37–47 (1997).

choice is instrumental and motivated in part by a belief that originalism tends to yield conservative results?[21] Second, even if we assume that he has given a complete account of his reasons for employing originalism, the fact that his choice of originalism is principled does not necessarily remove it from the realm of ideology: there is nothing contradictory about the notion of a principled ideologue. Third, the principles underlying his commitment to originalism might themselves be considered both "political" and "ideological." Why is Justice Scalia's conception of the proper role of judges in a democracy, and of the demands of the rule of law, not a "political" view? If he decides cases on the basis of his commitment to originalism, are we to say that his behavior is legally motivated, ideologically motivated, or both?

There are other ways in which judicial behavior may resist easy categorization as either "legal" or "ideological," such that one label applies to the exclusion of the other. Law and ideology are not, in fact, mutually exclusive categories: the "law" may explicitly give room for a judge's "ideology" to operate, while a judge's "ideology" may include a preference for following the law. Consider, for example, the fact that judges are frequently directed by law to exercise what is known as "discretion."[22] A grant of discretion implies that judges are free, within the legally defined bounds of their discretion, to reach varying results on the basis of whatever considerations they happen to deem relevant or appropriate. Thus, a judge who consistently exercises her discretion in a conservative direction is, in a sense, behaving ideologically, yet she is not behaving in a manner inconsistent with law.[23] Her behavior is both legally and ideologically driven: to the extent that the relevant legal criteria provide no further guidance, the law permits, if not invites, her to behave ideologically.

[21] *See* Sara C. Benesh & Jason J. Czarnezki, *The Ideology of Legal Interpretation*, 29 WASH. U. J.L. & POL'Y 113 (2009); Alexander Volokh, *Choosing Interpretive Methods: A Positive Theory of Judges and Everyone Else*, 83 N.Y.U. L. REV. 769, 805 (2008) (arguing that a "results-oriented judge" will select the interpretive theory that tends to produce the substantive results he or she prefers and that "observing, say, textualist decisions in the world may tell us more about *textualists* than it tells us about *textualism*").

[22] *See* Pauline T. Kim, *Lower Court Discretion*, 82 N.Y.U. L. REV. 383, 408–26 (2007).

[23] *See, e.g.*, POSNER, *supra* note 15, at 41–49; Kim, *supra* note 22, at 408–17.

To be sure, it is possible for judges to exercise discretion in a nonideological manner. But it is also possible for judges to exercise discretion in an ideological manner while remaining entirely faithful to law. Judges themselves find it uncontroversial that there are hard cases in which the law gives out, and in which they therefore can and do draw upon their personal views and preferences rather than choose arbitrarily.[24] They can and do disagree, of course, over the extent to which they have discretion, and that disagreement itself can be ideologically motivated. But to the extent that judges are operating within an area of true discretion, their reliance upon ideological considerations is neither contrary to, nor to the exclusion of, what the law demands.

Because the concept of "judicial ideology" relies upon an inherently murky and confused distinction between what is "legal" and what is "ideological" or "political," its use is bound to provoke disagreement between those who regard a particular form of judicial behavior to be "political" or "ideological" in character, and those who consider the same behavior to be "legal" in character. In cases of such disagreement, legal scholars can be expected to resist the use of the term "ideological" by political scientists to describe the behavior in question.

It is not possible for this Article to settle the correct meaning of the term "ideology." How researchers should or will use the term will inevitably depend upon the purposes that they have in mind. Our purpose here has been instead to highlight both the inherent difficulty of defining the term in a coherent and satisfying way, and the resulting risk that empirical research on the subject will be misunderstood or rejected. Different audiences may interpret the term

[24] *See, e.g.*, BENJAMIN N. CARDOZO, THE NATURE OF THE JUDICIAL PROCESS 16–17, 66 (1949) (arguing that the "method of free decision" is "dominant" in constitutional adjudication, and that judges are required to reach decisions that advance social welfare); POSNER, *supra* note 15, at 78–92, 102 (describing appellate judges as "occasional legislators" who are influenced by their "political leanings" when deciding "legalistically indeterminate" cases); Edwards, *supra* note 2, at 622 (acknowledging that "there are certain sorts of cases in which a judge's moral and political views unavoidably come into play"); Wald, *supra* note 10, at 250–51 ("[J]udges do have personal ideologies which sometimes enter into their decisionmaking. But how could it be otherwise? [W]hen a judge is presented with . . . a 'very hard' case . . . that judge must use her discretion to reach what she feels is the most appropriate and accurate result. [T]hat decision must necessarily be shaded by the judge's experience and beliefs.").

differently. It thus behooves scholars to specify as clearly as possible what they are choosing to study.

In practice, empirical researchers tend to use the term "ideology" to describe a judge's predisposition to decide certain types of cases in particular way. They do so on the implicit assumption that the basis of this predisposition is an "ideology" in the dictionary sense of the word—namely, a set of related ideas, preferences, or beliefs that are at least arguably political in character. As we discuss below, however, there is ordinarily no way for empirical scholars to know for certain whether this assumption holds true.[25] Our own use of the term "ideology" in this Article varies, as it must, with the context of the discussion. For the purpose of discussing in Part II why "ideology" is inherently difficult to measure, we use the term in its dictionary sense, which constitutes some kind of unattainable ideal. For the purposes of describing and evaluating in Parts III and IV the ways in which scholars have in fact sought to measure ideology, by contrast, we use the term in the less demanding sense that has been employed by other empirical scholars.

II. METHODOLOGICAL CHALLENGES

Problems of definition breed problems of measurement. We cannot know what data we must collect, and what methods we should apply, if we do not first specify what it is that we are attempting to measure. If scholars do not know what they mean by "ideology," they cannot hope to select the most appropriate way to measure it. Likewise, if scholars fail to specify what they mean by ideology, it becomes impossible for the audience to judge the appropriateness of the measures used.

The question of how judicial ideology should be defined can only be answered in light of why we care about judicial ideology in the first place and what our research goals happen to be. Is the goal, for example, to predict the likelihood that a given judge will arrive at decisions that are pleasing to conservatives? Are we interested in the

[25] *See infra* Parts II.A–B (discussing the inherent unobservability of a person's ideology, in its dictionary sense, and the simultaneous necessity and difficulty of drawing inferences about a judge's "ideology" from his or her observable behavior).

extent to which case outcomes depend on the identity of the judge or the composition of the panel? In either case, it may not be necessary to impute motivations to the judge at all. Is the goal instead to measure the ideology of the judge? If so, it will be necessary to define “ideology.” Or is the goal not only to measure the judge’s ideology, but also the extent to which that ideology influences the judge’s decision-making? If so, it will be necessary not only to define judicial ideology, but also to distinguish the effect of “ideological” factors from the effect of all “nonideological” factors. Is the goal to measure whether the judge is influenced *at all* by ideology? Or is the goal to measure the extent to which the judge is influenced by ideological considerations *to the detriment of legal considerations*? If it is the former, then it will be appropriate to analyze all cases in which the judge exercises discretion; if it is the latter, then it will be necessary to identify and analyze only those cases in which the judge exceeds the bounds of his or her discretion for ideological reasons. There are subtle, but important, substantive differences among these various research goals. Only after we have specified our research goals can we attempt to identify the definitional and measurement approaches that best advance those goals.

Even if one arrives at an explicit and goal-appropriate definition of “ideology,” however, obstacles to proper measurement abound. Three such obstacles will be discussed here: (1) the fact that judicial ideology is a latent trait that cannot be directly observed, with the result that we must rely upon inferences drawn from observable behaviors; (2) the problem of observational equivalence; and (3) the multidimensionality of judicial ideology.

A. The Inherent Unobservability of Ideology

What does it mean to say that ideology is a latent trait? Unlike a person’s age or sex, a person’s ideology cannot be directly observed. Indeed, it cannot be proven that “ideology” even exists in the form that scholars posit. The day may eventually arrive when it becomes possible, via functional magnetic resonance imaging or some analogous technology, to identify and label a phenomenon called “ideology” at work in the brain of a judge. Until that day arrives, however, the human mind remains largely a black box. Absent the ability to peer inside a judge's mind and observe a thing called "ideology" at work, the only way to measure "ideology" is to focus upon some observable trait or behavior that is correlated with, or indicative of, ideology.

Suppose, for example, that an empirical researcher wishes to measure David's liking for tofu. A taste for tofu is, like political ideology, a latent trait that cannot be directly observed. There are two approaches that the researcher might take to ascertain the extent to which he likes tofu. The first approach would be for the researcher to use some observable *trait* as a proxy for his unobservable attitude toward tofu. For example, the researcher might seize upon the fact that inhabitants of St. Louis consume less tofu per capita than inhabitants of Tokyo. His city of residence, unlike his state of mind toward tofu, is observable. Having observed that David lives in St. Louis and not Tokyo, the researcher might conclude that he does not like tofu.

The drawbacks to the use of this crude geographical proxy for food preference are obvious. Although the proxy may rest upon a generalization that tends on the whole to be true—namely, people in St. Louis consume relatively less tofu because they like it less—there is no guarantee that the generalization holds true in the specific case of David. Moreover, even if one observable characteristic indicates that he is not a likely tofu-lover, other observable traits—such as the facts that he is ethnically Asian and has lived in California—may suggest the opposite. Given a certain amount of information about both the shared observable traits of people who enjoy tofu and the extent to which David possesses those traits, the researcher's response might be to construct a model that employs a range of variables—including not only David's current place of residence, but also his ethnicity, his gender, his party affiliation, his age, his educational level, the number of years that he has lived in heavily urban areas with a high Asian population, and so forth—to predict the likelihood that he will like tofu. Ultimately, however, any conclusion that the researcher reaches concerning David's taste for tofu will be a prediction based on generalizations about the taste of other people who share certain characteristics with him.

The second approach that the researcher might take would be to look for observable *behavior* that is consistent or inconsistent with a taste for tofu. Self-reporting is one form of observable behavior: the researcher might simply ask David how he feels about tofu. But self-reporting is not necessarily a reliable source of information about his tastes. He might lie because he wants the researcher to believe that he is a healthy eater, or because he fears that his friends and neighbors will ridicule him for professing a penchant for tofu. Or he might simply lack the time or inclination to complete a questionnaire or interview on the subject of his dietary preferences.

Knowing this, the clever researcher might attempt to monitor his actual food-purchasing habits. She might keep track of how much tofu he buys at the supermarket, for example, by obtaining the data that the supermarket collects about his purchases as part of its customer loyalty program. Such an approach would allow her to observe not only how frequently David buys tofu from that supermarket, and how much, but also whether he buys more when it is on sale, whether there are other foods that he substitutes for tofu depending upon their prices relative to one another, and so forth. This approach of observing actual behavior that might be expected to reveal his attitude toward tofu has a great advantage over the use of proxy variables: the researcher can avoid relying upon generalizations about the dietary preferences of people who share certain characteristics with David. Such generalizations are bound to be inaccurate in many cases, and the researcher who relies upon them has no way of knowing whether someone happens to be an exception. A behavioral assessment of dietary preferences, based upon observation of actual choices that a person makes, promises to capture individual-level variation in a way that proxy measures such as place of residence and ethnicity cannot hope to do.

B. The Problem of Observational Equivalence

All other things being equal, individualized measurement of latent traits on the basis of observable behavior may be preferable to reliance upon inherently inaccurate proxy measures. Even this approach, however, is not foolproof. A major obstacle to the behavioral assessment of latent traits stems from the fact that, by definition, latent traits such as a preference for tofu or a conservative political ideology cannot be directly observed. Researchers can only draw inferences about their existence and magnitude from observable behavior, and often, there is more than one inference that can be drawn from the same act. Observable behavior is frequently open to multiple interpretations that are compatible with different explanations of the behavior in question.

Suppose, for example, that our intrepid researcher observes that David never includes tofu in his grocery purchases. It would be reasonable to infer from this behavior that he dislikes tofu. But there are other possible explanations. Perhaps he orders tofu in restaurants but does not cook it at home, or he prefers to buy his tofu from a specialty grocery store about which the researcher lacks information. Or perhaps his partner detests tofu, and he is wary of choosing foods that she will not want to share. Ideally, a researcher would acknowledge all plausible competing explanations for his behavior and test each of them. She might seek to determine whether his propensity to consume tofu increases when he is offered different brands in the supermarket, or when he is dining in the absence of his partner. By identifying, then controlling for, the effect of these variables, she would arrive at a better picture of his true appetite for tofu. An ideal research design might take the form of an experiment: the researcher might want to place him in a room and offer him pairwise choices between tofu and other types of food, as well as between different types of tofu.

The measurement of judicial ideology raises the same problems as the measurement of one's taste for tofu. In order to measure either latent trait, scholars must instead measure some observable behavior or trait that they believe is correlated with the existence of the trait that is truly of interest. But there are limits to what can be inferred about a person's attitudes from the behavior that we are capable of observing. In the real world—the world in which scholars of judicial behavior must operate—ideal research design is more dream than reality. Experiments, for the most part, are out of the question. One may attempt to observe judges in a laboratory setting, but it is not reasonable to expect them to behave as they would in the real world: their motivations and incentives in such a setting are not the same as they are in actual adjudication. At the same time, the constraints upon the data available for collection in the real world make it difficult, if not impossible, to isolate the effect of variables of interest, or to control for the effect of potentially confounding variables. For example, if our research goal is to isolate the impact of constitutional text on Justice Scalia's voting tendencies in First Amendment cases, we will never have the opportunity to observe him deciding a series of cases in which the text of the First Amendment varies while all other variables remain constant. As a result, the problem of potentially faulty inference from observable behavior cannot be eliminated.

The insight that observable behavior is open to multiple interpretations goes by different names. Lawrence Baum refers to the "theoretical ambiguity of behavior";[26] others speak of "observational equivalence"[27] or "behavioral equivalence."[28] Whatever name one attaches to the problem, it is endemic to the study of judicial ideology, and it greatly complicates efforts to evaluate competing hypotheses about the causes of judicial behavior. The ongoing debate over the relative merits of the legal, attitudinal, and strategic models[29] persists in part because the behavior that we observe can often be explained in more than one way. Consider, for example, the fact that Chief Justice Rehnquist authored the majority opinion in *Dickerson v. United States*,[30] upholding *Miranda v. Arizona*,[31] a landmark liberal decision of the Warren Court. His authorship of *Dickerson* is consistent with several different explanations. One possibility is that he sincerely preferred to uphold *Miranda*. Such a liberal policy preference would have been highly uncharacteristic of him, however, in light of the highly conservative leanings he had previously demonstrated over the course of decades on the Court.[32] A second

[26] LAWRENCE BAUM, THE PUZZLE OF JUDICIAL BEHAVIOR 20 (1997).

[27] *E.g.*, Gary J. Miller, *The Political Evolution of Principal-Agent Models*, 8 ANN. REV. POL. SCI. 203, 210, 212, 222 (2005) (citation omitted); Brian R. Sala & James F. Spriggs, II, *Designing Tests of the Supreme Court and the Separation of Powers*, 57 POL. RES. Q. 197, 201–02, 204 (2004) (devising a test of the separation-of-powers model that navigates around the observational equivalence of attitudinal and strategic behavior).

[28] Kim, *supra* note 22, at 427.

[29] *See supra* note 15 and accompanying text.

[30] 530 U.S. 428 (2000).

[31] 384 U.S. 436 (1966).

[32] *See, e.g.*, Andrew D. Martin & Kevin M. Quinn, *Dynamic Ideal Point Estimation via Markov Chain Monte Carlo for the U.S. Supreme Court, 1953–1999*, 10 POL. ANALYSIS 134, 146 tbl.1 (2002) (identifying Chief Justice Rehnquist as the second-most conservative Justice to

possibility is that he voted against his ideological preferences: he might have preferred to strike down *Miranda* but felt constrained by legal considerations, especially the notion of stare decisis, to vote the other way. A third possibility is that he behaved both ideologically and strategically. Had he chosen to dissent, there would still have been six votes to uphold *Miranda*, and responsibility for assigning the task of writing the majority opinion would have fallen in that case upon Justice Stevens, the most senior member of the majority and also the most liberal member of the Court,[33] who might have seized for himself the opportunity to author a bolder and more expansive opinion. Nor do the legal, attitudinal, and strategic models exhaust the possible explanations for his vote. He may have been motivated by a desire for the esteem of a particular audience—say, the readers of the *New York Times*,[34] as opposed to the members of the Federalist Society.[35]

The judicial treatment of precedent offers another demonstration of the difficulty of inferring ideological motivations from observable behavior. Suppose that a judge dissents from the decision of the court in case A and later votes in case B to limit the precedential reach of case A. Her vote in case B could reasonably be interpreted as an ideologically motivated effort to undermine a decision that she had opposed from the outset on ideological grounds.[36] The fact that she votes merely to limit, rather than to overrule, case A can easily be explained as strategic behavior: her best strategy for achieving her ideological goals may be to build the necessary support for a formal overruling of case A over time, or simply to achieve the de facto overruling of case A by ensuring that it is gradually confined to its

serve on the Court over the period spanning from 1953 to 1999); SEGAL & SPAETH, *supra* note 15, at 322 tbl.8.2.

[33] *See* Martin & Quinn, *supra* note 32, at 146 tbl.1.

[34] *See* LAWRENCE BAUM, JUDGES AND THEIR AUDIENCES: A PERSPECTIVE ON JUDICIAL BEHAVIOR 139 (2006) (noting the use of the term "Greenhouse effect," named after *New York Times* reporter Linda Greenhouse, to describe the extent to which judges may be influenced by the promise of favorable media coverage).

[35] *See id.* at 123–26; BAUM, *supra* note 26, at 40.

[36] *See* SEGAL & SPAETH, *supra* note 15, at 77 ("[T]he justices have rarely acceded to [precedents] of which they disapprove."); *id.* at 81 (arguing that "precedent . . . provides virtually no guide to the justices' decisions" and "is a matter of good form, rather than a limit on the operation of judicial policy preferences"); *id.* at 288–311 (reporting that liberal Justices tend to uphold liberal precedents and to limit or overrule conservative precedents).

facts. But one might also reasonably interpret her behavior as motivated by legal, as opposed to ideological, considerations. The fact that her vote in case B is merely to limit, and not to overturn, the ruling in case A could reflect the constraining effect of precedent on her ideological motivations: given that it remains open to her to vote in favor of overruling case A, one might conclude that her behavior is guided not by ideological preference, but rather by the legal principle of stare decisis.[37]

The phenomenon of panel composition effects poses a number of related methodological challenges, among them the problem of observational equivalence. Over a decade ago, Professor Revesz[38] and Professors Cross and Tiller[39] discovered that the voting behavior of federal appeals court judges tends to vary with the partisan composition of the panels on which they happen to sit.[40] On a three-judge panel, a Democratic appointee tends to vote more liberally if paired with at least one other Democratic appointee than if he or she is the lone Democratic appointee, and to vote even more liberally if all three members of the panel are Democratic appointees; likewise, Republican appointees tend to vote more conservatively when they are in the majority than when they find themselves in the minority, and to vote even more conservatively when there is no Democratic appointee present at all.[41] One challenge that empirical scholars must address, therefore, is the fact that panel composition effects can conceal the true extent of a judge's ideological leanings. Because the influence of ideology on a judge's voting behavior may be muted unless he or she is paired with at least one likeminded colleague, a simple analysis of individual judicial voting records that fails to

[37] *See* Donald R. Songer & Stefanie A. Lindquist, *Not the Whole Story: The Impact of Justices' Values on Supreme Court Decision Making*, 40 AM. J. POL. SCI. 1049, 1051–52 (1996).

[38] *See* Richard L. Revesz, *Environmental Regulation, Ideology, and the D.C. Circuit*, 83 VA. L. REV. 1717 (1997).

[39] *See* 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).

[40] *See* POSNER, *supra* note 15, at 31–34 (reviewing different explanations that have been offered for the existence of panel-composition effects); Cross & Tiller, *supra* note 39, at 2171–76; Revesz, *supra* note 38, at 1751–56.

[41] *See* Cross & Tiller, *supra* note 39, at 2171–76; Revesz, *supra* note 38, at 1751–56.

control for panel composition is likely to underestimate the true extent of the judge's ideological preferences.

The other challenge that scholars face, however, is that of explaining why panel composition effects exist at all. The finding that judges tend to vote differently depending upon the partisan composition of the panel is open to a variety of explanations. One is the "whistleblower" hypothesis: on this view, the minority judge moderates the behavior of the other judges by threatening to expose "manipulation or disregard of the applicable legal doctrine."[42] A second explanation is the "dissent hypothesis": on this view, the judges moderate their positions in order to avoid the costs involved in writing and responding to a dissent.[43] A third explanation is the "deliberation hypothesis": on this view, the judges on an ideologically divided panel converge in their views as a result of substantive deliberation.[44] All three theories predict that judges on homogenous panels will show stronger ideological voting tendencies than judges on heterogeneous panels. If, however, the only behavior we ever observe is consistent with all three theories, then we have no way of ruling out any of the theories.

C. The Multidimensionality of Judicial Ideology

It is common to characterize judges and their votes as "conservative" or "liberal," in much the same manner as we might describe the attitudes and behaviors of any ordinary person or political actor.[45] Yet we also know that the terms "liberal" and "conservative" conceal a certain amount of heterogeneity. A judge may cast "conservative" votes in one category of cases (say, abortion) and "liberal" votes in another category (say, asylum). Whether we conclude that the judge in question is "liberal" or "conservative" will depend in such situations on which set of votes we happen to observe. By observing only one set of votes, we would

[42] Cross & Tiller, *supra* note 39, at 2156, 2171.

[43] *See* Revesz, *supra* note 38, at 1733–34.

[44] *See id.* at 1732–34.

[45] *See, e.g.*, SEGAL & SPAETH, *supra* note 15, at 86, 309–10, 329–31.

measure only one dimension of the judge's multidimensional ideology.

In practice, however, the challenge that multidimensionality poses to the measurement of judicial ideology may not be as severe as this hypothetical example might suggest. The views that people hold across a range of questions tend to correlate with one another in systematic ways. Electoral competition encourages the formation of political coalitions and the articulation of competing ideologies that distinguish one coalition from another and define the battle lines of popular politics. Indeed, it is the *systematic* correlation of views across related but distinct topics that defines an "ideology," or *system* of ideas. If political views were not correlated with one another, it would be impossible to speak of ideology at all; we would confront only disorganized collections of views lacking any kind of internal coherence at every turn. Nevertheless, even if it happens to be the case that, on the whole, judges who reach stereotypically conservative conclusions in abortion cases also tend to reach stereotypically conservative conclusions in asylum cases, generalizations from a judge's attitudes in one area of law to those in another may not always hold true.

It is difficult to identify actual settings in which judicial ideology is multidimensional because there is no generally accepted methodology for assessing the dimensionality of an ideology space.[46] There are a variety of scaling techniques available for estimating multidimensional voter preferences, but these techniques may yield disparate estimates of dimensionality.[47] Thus, it may not be possible

[46] *See* KEITH T. POOLE, SPATIAL MODELS OF PARLIAMENTARY VOTING 141–47 (2005); James J. Heckman & James M. Snyder, Jr., *Linear Probability Models of the Demand for Attributes with an Empirical Application to Estimating the Preferences of Legislators*, 28 RAND J. ECON. S142, S170–S171 (1997).

[47] *Compare, e.g.*, KEITH T. POOLE & HOWARD ROSENTHAL, CONGRESS: A POLITICAL-ECONOMIC HISTORY OF ROLL CALL VOTING 27–29 (1997) (arguing that only two dimensions are necessary to represent congressional preferences), *with* Heckman & Snyder, *supra* note 46, at S171–S173 (arguing that at least six dimensions are necessary). *See also* Timothy J. Brazill & Bernard Grofman, *Factor Analysis Versus Multi-Dimensional Scaling: Binary Choice Roll-Call Voting and the U.S. Supreme Court*, 24 SOCIAL NETWORKS 201 (2002) (showing that factor analysis yields systematically higher estimates of dimensionality than multi-dimensional scaling as applied to both simulated data and actual Supreme Court voting data).

to verify the adequacy of a one-dimensional model of judicial ideology using current methods.

Recent research provides some support for the assumption of unidimensional ideology that underlies most empirical scholarship on the U.S. Supreme Court. A study by Professors Grofman and Brazill estimates that between 80% and 93% of the variation in the Justices' voting patterns over the period from 1953 to 1991 can be explained in terms of disagreement along a single dimension.[48] That study reports, moreover, that the extent to which voting on the Supreme Court is unidimensional appears only to be increasing over time.[49] In a different paper, however, Grofman and Brazill observe that their methodology appears to overstate the importance of the primary dimension.[50] Although the full extent of the problem is unknown, their results show that at least 14% of the variation in Supreme Court voting *cannot* be explained by a single dimension of disagreement.[51] This shortfall may be enough to render the unidimensionality assumption problematic for many applications.

At the same time, the assumption of unidimensional ideology may not hold as well for other courts as it does for the United States Supreme Court. Very little is known about the dimensionality of ideology on other courts, and further investigation is surely warranted. Recent studies, for example, suggest that ideological disagreement in the Supreme Court of Canada may be significantly more multidimensional than in the U.S. Supreme Court.[52] The voting

[48] *See* Bernard Grofman & Timothy J. Brazill, *Identifying the Median Justice on the Supreme Court Through Multidimensional Scaling: Analysis of "Natural Courts" 1953–1991*, 112 PUB. CHOICE 55, 58 (2002). The 80% figure is a conservative estimate; the amount of variation depended on the model used. *See id.* (reporting that unidimensional models explained from 80% to 93% of the variation).

[49] *See id.* (observing that "by the time we get to the Rehnquist courts the finding of strong unidimensionality is indisputable").

[50] *See* Brazill & Grofman, *supra* note 47, at 217.

[51] *See id.* (reporting that, on average, a unidimensional approach to the Supreme Court voting data analyzed by the authors explained 86% of the variance).

[52] *See* Benjamin R.D. Alarie & Andrew Green, *The Reasonable Justice: An Empirical Analysis of Frank Iacobucci's Career on the Supreme Court of Canada*, 57 U. TORONTO L.J. 195 (2007); C.L. Ostberg et al., *Attitudinal Dimensions of Supreme Court Decision Making in Canada: The Lamer Court, 1991–1995*, 55 POL. RES. Q. 235, 242–49 (2002) (concluding on the basis of factor analysis that ideological disagreement on the Canadian Supreme Court occurs along three dimensions-namely, a communitarian versus libertarian dimension, a "fairness"

behavior of Justice Claire L'Heureux-Dubé illustrates the problem vividly. In their analysis of all votes cast on the Canadian Supreme Court over a period of more than thirteen years, Professors Alarie and Green report that she compiled the most "conservative" voting record of anyone on the court;[53] indeed, their estimate of her ideal point identifies her as an "extreme" justice.[54] They are quick to note, however, that their overall characterization of Justice L'Heureux-Dubé is confounded by her radically divergent voting patterns across different areas of law. In criminal appeals and cases involving claims under the Canadian Charter of Rights and Freedoms, her voting record was indeed the most conservative, and significantly more conservative than that of the median justice.[55] In labor cases, however, she compiled the most *liberal* voting record of the sixteen justices included in the study, while in the area of aboriginal law, hers was the fourth-most liberal voting record.[56]

As Alarie and Green freely acknowledge, their ideal-point estimation techniques—which are the same as those underlying the influential Martin-Quinn scores for Supreme Court Justices in the United States[57]—rest upon the assumption that the ideological ordering of the justices "should not depend on the underlying area of law,"[58] but this assumption proves highly problematic in the case of Justice L'Heureux-Dubé. Her overall characterization as "conservative" and "extreme"[59] reflects as much the happenstance composition of the court's docket as her own behavior: if the docket were sufficiently dominated by labor and aboriginal cases, the same estimation techniques would identify her as liberal rather than conservative. Nor is there any obvious way to correct the weighting

dimension in the area of criminal procedure, and a "judicial activism" versus self-restraint dimension).

[53] *See* Alarie & Green, *supra* note 52, at tbl.1, 209–14, 209 fig.2, 210 tbl.2.

[54] *Id.* at 209–10.

[55] *See id.* at 210.

[56] *See id.* at 207 tbl.1, 210; *see also* Ostberg et al., *supra* note 52, at 242–44 (characterizing Justice L'Heureux-Dubé as having a "communitarian orientation" that is more conservative on issues of criminal justice but more liberal with respect to the treatment of disadvantaged groups).

[57] *See id.* at 205–07 (citing Martin & Quinn, *supra* note 32).

[58] *Id.* at 210.

[59] *See supra* notes 53–54.

of different types of cases in order to produce an undistorted picture of her ideology: any weighting that researchers might devise runs the risk of appearing arbitrary.[60]

III. Methods for Measuring Judicial Ideology

Statistical analysis has become so commonplace in the study of judicial behavior that it is easy to overlook the steps that are necessary to convert cases, judges, and decisions into numerical quantities to be analyzed. A variety of methods have been developed for quantifying decisions, judicial ideology, and characteristics of cases. There is no "ideal" method for quantifying cases; each of these methods has its merits and its shortcomings. Too often, however, little attention is paid to these crucial methodological choices, and methods of coding cases or measuring ideology seem to be chosen casually based on "current practice," without careful attention to the objectives of the study. As we will demonstrate, the way that case outcomes and ideology are measured can have a large impact on the results of a study. Careful consideration of how cases are selected and outcomes are coded is necessary if we are to interpret the results of any study properly, or to understand what hypotheses can, or cannot, be tested within a particular empirical framework.

Part III.A looks at the first step of the quantification of cases: the coding of decisions. This may be done by reference to the presence or absence of particular actions, such as overturning agency rulings, or providing relief for certain kinds of plaintiffs. Alternatively, judicial actions may be labeled subjectively by the researcher as "liberal" or "conservative." Finally, studies may employ an "agnostic" coding scheme, by which the ideological direction of a particular outcome is a function of the voting alignment of judges in the case. Any of these coding methods can be used in a narrow data set focusing on a particular class of cases, or on a broad aggregation of cases across varying issue areas.

[60] *See* David S. Law, *Globalization and the Future of Constitutional Rights*, 102 NW. U. L. REV. 1277, 1300–01 n.104 (2008) (discussing the problems inherent in constructing any index that weights and combines different dimensions of a given phenomenon).

Part III.B discusses different ways of measuring judicial "ideology." Empirical scholars on courts have developed several methodologies for surmounting the various obstacles to the measurement of judicial ideology described in Parts I and II. One method, used frequently for the courts of appeals and district courts, is to use "proxy variables"—variables constructed from observable characteristics of judges that are believed to be correlated with their true ideologies. Such proxy variables have traditionally included the party of the President who appointed the judge, or the judge's race, gender, or professional background. More recently, more sophisticated "composite proxy variables" have been developed that attempt to account for the influence of competing actors over the nomination and confirmation of judges.[61]

An alternative—and often preferable—approach is to treat each judge's ideology as a latent variable to be estimated from data on the judge's actual behavior. These behavioral-assessment methods for estimating ideology range quite widely in sophistication, from simple vote-counting approaches to extraordinarily complex structural models. The advantage of measuring ideology in this manner is that it avoids the measurement error that is intrinsic to all proxy variables.

Finally, we describe a method that might best be thought of as a hybrid of the two previous methods. In this approach, estimates of judicial ideology derived from data on a judge's actual behavior in one context are "transplanted" into a different context. Although this approach has the potential to combine the best of both worlds by marrying the precision of behavioral assessment with the convenience of proxy variables, the results produced in this manner can be difficult to interpret and require extreme care. We will discuss some of the benefits and pitfalls of these methods below.

[61] *See, e.g.*, Micheal W. Giles et al., *Picking Federal Judges: A Note on Policy and Partisan Selection Agendas*, 54 POL. RES. Q. 623 (2001); Jeffrey A. Segal & Albert D. Cover, *Ideological Values and the Votes of U.S. Supreme Court Justices*, 83 AM. POL. SCI. REV. 557 (1989).

A. Ways of Coding Cases

One of the most challenging aspects of the quantitative study of judicial decision-making is that there is nothing inherently quantitative about a judicial decision.[62] It is therefore necessary for researchers to select and employ some method for converting judicial decisions into numerical quantities for statistical analysis. Each of the most common methods for doing so, however, embodies different assumptions that can have considerable implications for the meaning and interpretation of a study's findings.

Most empirical studies on the subject of judicial ideology rely on some sort of dichotomous coding scheme, in which observable judicial actions–typically a vote or decision of some kind–are coded as "zero" or "one," depending on whether they are, in some rough sense, "liberal" or "conservative." Often, regression analysis is then used to estimate the impact of some measure of ideology on the actions that have been coded in this manner. It is impossible to interpret the results of this type of analysis correctly, however, unless we first understand precisely why a given vote or outcome has been characterized as "liberal" or "conservative." As we explore below, each of the common methods of coding cases embodies a particular conception of judicial ideology.

There are three ways in which researchers commonly code cases, which correspond to three distinct conceptions of ideology. First, the researcher can implement a "narrow" approach to coding, and thus a "narrow" conception of judicial ideology as well, by measuring judges' propensities to rule in a particular direction in a certain area of law or on a specific legal issue. Second, the researcher can adopt a "broad approach" by measuring the propensity of judges to rule in a "liberal" or "conservative" direction, as defined subjectively by the researcher, across a broad range of cases. Third, the researcher can pursue an "agnostic" approach by measuring the propensity of judges to vote in the same way as certain other prespecified judges. Unlike the "narrow" approach, the "agnostic" approach can be applied across a range of cases. Unlike the "broad" approach, however, the

[62] Exceptions include sentencing decisions and damage awards.

"agnostic" approach does not require the researcher to apply value labels to votes or outcomes. The least intuitive of the three approaches (at least at first glance), it is agnostic in the sense that the researcher does not need to decide whether any particular outcome is "liberal" or "conservative," but rather allows the directionality of each outcome to be determined by the voting alignments among the judges.

1. The "Narrow" Approach

The most common way to code outcomes, which implements what we call a "narrow" conception of ideology, is to examine cases in a single area of law, or that address a particular legal issue, and to designate cases as "pro-plaintiff" if the rulings provide some threshold level of relief to the plaintiff. Depending on the context, readers may recognize these outcomes as "liberal" or "conservative"; for example, a ruling in favor of a union in a labor case or a defendant in a criminal case would commonly be referred to as "liberal." But when the data is limited to a particular issue area, these terms should be construed narrowly.

The advantage to limiting inquiry to a single issue area (or analyzing multiple issue areas separately) is that judicial preferences are likely to be unidimensional *within a given issue area*. Judges may well hold preferences that are unidimensional across the entire range of cases that they decide, such that a judge who is liberal in criminal cases tends also to be liberal in sex discrimination cases, and vice versa. As discussed above in Part II, however, this assumption may not always hold true. Confining the analysis to a single issue area makes it unnecessary to assume that a judge who is more "liberal" on a particular issue must prefer "liberal" outcomes over other unrelated issues. Nor does one need to identify which outcome is "liberal" in ideologically cross-cutting areas of law. It is simply enough to assume that some judges are more or less likely to support asylum claims, or employment discrimination claims, or industry challenges of agency rulemaking. To the extent that judicial ideology is multi-dimensional, a narrow, issue-specific approach to the measurement of ideology is appropriate because it avoids the conflation of dissimilar preferences or, in more colloquial terms, the comparison of apples and oranges. Proper delineation of a specific issue area then becomes crucial if we are to maintain a plausible assumption of one-dimensional preferences.

An issue-specific approach to the measurement of judicial ideology is not foolproof. Delineation of the issue area can itself pose challenges, as the contours of the area may vary depending on the court, the time period, and the type of cases presented. Cases that present multiple issues may defy easy categorization and require scholars to exercise judgment in deciding which of them should be included in a particular study.[63] Even in what is commonly understood to be a specific area of law—say, criminal law—the data may need to be subdivided further in order to ensure that apples are compared only to apples. Depending upon the context, it may be proper to separate these cases into those involving white-collar and blue-collar crimes, due process claims, sentencing challenges, issues of statutory interpretation, and so forth. Moreover, even if one addresses the problem of multidimensionality by measuring ideology on an issue-by-issue basis, the problem of observational equivalence remains: it may only be possible to measure judges' *propensities* to vote in a particular direction, and not necessarily their *motivations*. A tendency to favor asylum claimants, for instance, could stem from sympathetic attitudes toward immigrants, or from a preference for a particular interpretation of the relevant asylum statutes, or from lack of deference toward agency adjudications in general or immigration courts in particular.

2. The "Broad" Approach

The other common approach to case-coding is to collect and aggregate data that span many issue areas, on the premise that a single dimension explains judges' preferences across these many

[63] *See* Paul H. Edelman & Jim Chen, *The Most Dangerous Justice Rides Into the Sunset*, 24 CONST. COMMENT 299, 305–08 (2007) (arguing that the manner in which Harold Spaeth's widely used Supreme Court database is coded ignores the multiplicity of issues present in most Supreme Court cases); Carolyn Shapiro, *Coding Complexity: Bringing Law to the Empirical Analysis of the Supreme Court*, 60 HASTINGS L.J. 477, 488–530 (2009) (discussing, and attempting to measure, the extent to which the manner in which the coding of Spaeth's database masks important information about the legal issues present in each case).

dimensions. This approach is premised upon what we might call the "broad" conception of ideology. It is commonly used in studies of the Supreme Court; because the Court hears relatively few cases, it may not be feasible to separate cases by issue and still have enough data for meaningful empirical analysis. Furthermore, as discussed in Part II.C, a single ideological dimension explains a large part of Supreme Court voting. In such studies, the ideology spectrum is necessarily amorphous: it could represent a preference for certain kinds of policy outcomes, or for methods of statutory or constitutional interpretation, or any combination of these and other influences.

Whether a unidimensional approach to the measurement of judicial ideology across different issue areas is appropriate will depend upon the application. Although a single dimension explains much of the voting behavior on the Supreme Court, this varies substantially by issue area, as we demonstrate in Part IV.B. A single dimension, for example, explains a large component of voting in criminal procedure and civil liberties cases, but is much less effective at explaining economic and tax cases. If the primary goal of the research is to derive a rough estimate for the influence of ideology in a particular court, or to examine how the ideology of the Justices changes over time,[64] then a single dimension may well be sufficient (provided, of course, that one can arrive at a substantively satisfying definition of ideology in the first place). Likewise, the unidimensional approach may suffice if the goal is to study the effect of external influences on judicial behavior, for example, such as the extent to which Congress constrains the Supreme Court[65] or appellate review constrains lower courts.[66] However, it may be wise to restrict

[64] *See, e.g.*, Martin & Quinn, *supra* note 32; Andrew D. Martin & Kevin M. Quinn, *Assessing Preference Change on the US Supreme Court*, 23 J.L. ECON & ORG. 365, 366 (2007); Lee Epstein et al., *Ideological Drift Among Supreme Court Justices: Who, When, and How Important?*, 101 NW. U. L. REV. 1483, 1493–97 (2007).

[65] *See, e.g.*, Mario Bergara et al., *Modeling Supreme Court Strategic Decision Making: The Congressional Constraint*, 28 LEGIS. STUD. Q. 247 (2003); Anna Harvey & Barry Friedman, *Pulling Punches: Congressional Constraints on the Supreme Court's Constitutional Rulings, 1987–2000*, 31 LEGIS. STUD. Q. 533 (2006); Jeffrey A. Segal, *Separation-of-Powers Games in the Positive Theory of Congress and Courts*, 91 AM. POL. SCI. REV. 28 (1997); Pablo T. Spiller & Rafael Gely, *Congressional Control or Judicial Independence: The Determinants of U.S. Supreme Court Labor-Relations Decisions, 1949–1988*, 23 RAND J. ECON. 463 (1992).

[66] *See, e.g.*, David E. Klein & Robert J. Hume, *Fear of Reversal as an Explanation of Lower Court Compliance*, 37 LAW & SOC'Y REV. 579 (2003); Donald R. Songer et al., *The*

analysis to issue areas that are well explained by the unidimensional approach.

To their disadvantage, however, unidimensional models that aggregate cases over multiple issues compound the problem of observational equivalence. Those using the aggregative approach must deal not only with the fact that cross-cutting motivations may be at work in the same issue area, but also with the additional challenge of determining the extent to which a given motivation applies across different areas. If, for example, both originalism and political conservatism could lead to the same combination of judicial votes, it may be impossible to distinguish between these two theories of behavior in a single area of law, much less across multiple areas. Such theories can only be tested by a model that allows preferences along multiple dimensions, in the context of cases that generate a clash between the two motivations.

A final difficulty with aggregating cases along a single dimension is that it may be impossible to determine the direction of the outcome objectively. If a state is defending a liberal policy in a federalism case—say, *Gonzales v. Raich*[67]—what position is to be considered the "liberal" one? Traditionally, liberal judges have been advocates of federal power in federalism disputes, but one could imagine, in a close case, that liberal judges might be more likely to uphold the state's liberal policy against a claim of federal power. Such a case presents a tension between preferences along two dimensions—federalism and social policy—that can only be resolved on the basis of some explicit understanding of the relative importance of the two dimensions to the case and to the judges themselves.

The potential perils of unidimensional case coding are highlighted by two recent working papers[68] that criticize the subjective coding

Hierarchy of Justice: Testing a Principal-Agent Model of Supreme Court-Circuit Court Interactions, 38 AM. J. POL. SCI. 673 (1994).

[67] 545 U.S. 1 (2005) (ruling, in favor of the federal government, that California lacked the power to legalize the use of marijuana for medicinal purposes).

[68] *See* Anna Harvey, What Makes a Judgment "Liberal"? Coding Bias in the United States Supreme Court Judicial Database (June 15, 2008), http://politics.as.nyu.edu/docs/IO/2787/harveymeasurementerror.pdf; William M. Landes & Richard A. Posner, *Rational Judicial Behavior: A Statistical Study*, 1 J. LEGAL ANALYSIS (forthcoming 2009), *available at* http://ssrn.com/abstract=1126403.

decisions found in the most widely used database of Supreme Court decisions, Harold Spaeth's U.S. Supreme Court Judicial Database.[69] In one, Professor Harvey considers cases that involve constitutional challenges to federal statutes and estimates the "ideology" of a statute by using a measure based on the members of Congress who voted in favor of it.[70] For these cases, the correlation between the objective measure and the subjective coding in the Spaeth database is low, and the association is statistically insignificant.[71] Harvey's findings suggest that many cases may not be amenable to coding on a single "liberal" to "conservative" dimension, and that the results of studies that do rely upon such unidimensional coding schemes may well be artifacts of subjective coding decisions.

In the other paper, Professor Landes and Judge Posner reanalyze the coding decisions made in Spaeth's Supreme Court database and Donald Songer's Court of Appeals database[72] and challenge the original coding choices in many case categories.[73] With respect to the Supreme Court database, Landes and Posner take issue with the coding decisions made in such areas as commercial speech, campaign finance, labor law, and administrative law.[74] With respect to the Court of Appeals database, they object to the coding choices made in a variety of legal contexts including white-collar crime, same-sex harassment, and intellectual property.[75] Many of these categories are indeed difficult to code because they present conflicting issues. Without taking a position on the proper way to code outcomes in any

[69] *See* Supreme Court Data, http://www.cas.sc.edu/poli/juri/sctdata.htm (follow "The Original U.S. Supreme Court Judicial Database" hyperlinks for documentation and data in different formats) (last visited Apr. 14, 2009).

[70] The measure is the change in status quo of the Poole-Rosenthal "Nominate" measures of the statute, based on the roll call of Senators and Representatives who voted for and against it. *See* POOLE & ROSENTHAL, *supra* note 47.

[71] *See* Harvey, *supra* note 68, at 19–20. The correlation is 0.12. E-mail from Anna Harvey, Associate Professor of Politics, New York University, to Joshua Fischman, Assistant Professor of Economics, Tufts University (Mar. 28, 2008, 10:26:12 EDT) (on file with the authors).

[72] *See* Appeals Court Data, http://www.cas.sc.edu/poli/juri/appctdata.htm (follow hyperlinks to different versions of the data and corresponding documentation) (last visited Apr. 14, 2009).

[73] *See* Landes & Posner, *supra* note 68.

[74] *Id.* at 42.

[75] *Id.* at 43–44.

of these categories, it suffices to note that highly distinguished scholars can disagree on the directionality of Supreme Court rulings. Whether the outcomes of any studies hinge on their particular coding decisions remains to be seen. But this disagreement suggests that it may be impossible to construct an objective measure of coding Supreme Court decisions in a single dimension.

3. The "Agnostic" Approach

In order to avoid these difficulties, some studies employ an "agnostic" method of coding cases that does not require the researcher to make a subjective assessment of the direction of each outcome. Rather, these coding schemes *estimate* the direction of the outcome from the voting alignment of the judges: if the "liberal" judges all vote in favor of the same outcome, that must be the "liberal" outcome. Agnostic coding can be counterintuitive because the ideology of the judges must be inferred from their votes at the same time that the directionality of each case is inferred from the judges' positions. On first impression, the process may seem to involve a degree of circularity or bootstrapping: how can the direction of case outcomes and the ideology of judges be inferred simultaneously from one another? The key to understanding how agnostic coding models work is to realize that they do not measure ideology with reference to any particular kind of concrete outcome; rather, they measure ideology purely in terms of voting alignments.

Interpretation of the results from such models requires an understanding of the scale along which judicial ideology is being measured. The typical way of defining the scale is to identify the judges who represent the "liberal" and "conservative" ends of the spectrum, and to treat them as anchor points. Consider the most well-known application of agnostic coding: the "ideal points" that Professors Martin and Quinn calculate for Supreme Court Justices.[76] In this context, to be "liberal" might mean to be "Brennan-like": a "liberal" Justice is one who often voted with Brennan, or who often voted with Justices who often voted with Brennan. Similarly, to be

[76] *See* Martin & Quinn, *supra* note 32, at 147–51.

"conservative" might mean to be "Rehnquist-like."[77] For an American audience, this interpretation is quite intuitive; most readers are familiar with the positions that liberal and conservative Supreme Court Justices typically take. Nevertheless, the interpretation of this agnostic ideology spectrum requires some understanding of the Justices' positions that is external to the study. Consider, for example, Alarie and Green's estimation of an agnostic model of ideology on the Canadian Supreme Court.[78] One could conceive of the ideological spectrum of Canadian justices as ranging from "L'Heureux-Dubé-like" to "Sopinka-like."[79] However, for readers who are unfamiliar with Canadian constitutional law—and, indeed, perhaps even for Canadian constitutional scholars—such an ideological scale may prove less than intuitive.

Although agnostic coding methods are often associated with highly sophisticated empirical studies, such as the Martin-Quinn approach to ideal point estimation,[80] they have also found use in many non-technical applications. In some of the earliest empirical studies of the Supreme Court, the political scientist Herman Pritchett constructed tables showing how often each pair of Justices was in agreement, or how often each pair dissented together, within a particular time period.[81] Pritchett's approach constituted a form of agnostic coding because he only examined whether Justices voted together, and did not make any effort to label individual votes as "liberal" or "conservative." Using these indicators of agreement, Pritchett was able to identify blocs of Justices who often voted together. For example, he concluded that in the 1939 and 1940 Terms, there were two voting blocs—one consisting of Justices McReynolds, Roberts, Hughes, and Stone, and another consisting of Justices Frankfurter, Murphy, Douglas, and Black—with Justice

[77] The actual Martin-Quinn model uses more than two anchor points. It includes prior assumptions about the starting ideal points of Justices Harlan, Douglas, Marshall, Brennan, Frankfurter, Fortas, Rehnquist, Scalia, and Thomas. *See id.* at 147.

[78] Alarie & Green, *supra* note 52; *supra* notes 52–56 and accompanying text.

[79] *See* Alarie & Green, *supra* note 52, at 210 tbl.2.

[80] *See* Martin & Quinn, *supra* note 32, at 147–51; *supra* text accompanying note 76.

[81] *See, e.g.*, C. Herman Pritchett, *Divisions of Opinion Among Justices of the U. S. Supreme Court, 1939–1941*, 35 AM. POL. SCI. REV. 890 (1941) [hereinafter Pritchett, *Divisions of Opinion*]; C. Herman Pritchett, *The Roosevelt Court: Votes and Values*, 42 AM. POL. SCI. REV. 53 (1948).

Reed in the middle as the swing vote.[82] Pritchett identified the former bloc as the "right wing" and the latter bloc as the "left wing,"[83] but these labels required some familiarity with the Justices and the issues of the time that was external to his empirical analysis.[84]

Agnostic coding holds powerful attractions for empirical scholars of judicial behavior. Because the ideological direction of each outcome is inferred from the data, there is generally little need for concern that the results reflect subjective coding decisions by the author of the study.[85] Another advantage of agnostic coding is that it facilitates the analysis of large data sets by reducing the amount of effort required to code each case. In a recent paper, for example, Professor Ho was able to analyze over 17,000 FCC adjudications over a period spanning four decades.[86] This already formidable project might have been prohibitively impractical, had a subjective assessment of ideological orientation been necessary for each adjudication.

But the agnostic approach to coding has its shortcomings as well. For starters, models of this type generally assume a one-dimensional spectrum, yet it may be difficult to know what distortions this assumption might cause. Although methods exist for estimating the dimensionality of ideology on multimember courts[87]—at least when the composition of the court is fixed and judges vote sincerely—these methods are not applicable to courts in which judges decide cases alone or on rotating panels. If the assumption of unidimensional

[82] *See* Pritchett, *Divisions of Opinion*, *supra* note 81, at 893–95.

[83] *See id.* at 895.

[84] *See* G. Edward White, *Unpacking the Idea of the Judicial Center*, 83 N.C. L. REV. 1089, 1101–04 (2005).

[85] As a practical matter, however, agnostic coding and the more traditional coding approach used in the popular Spaeth database, *see supra* notes 63, 68–75 and accompanying text, may in practice yield highly similar results, at least in the context of the U.S. Supreme Court. When Professor Bafumi and his colleagues estimated an agnostic model of Supreme Court voting, they found that the direction of the outcome predicted by their model coincided with the coding of such votes in the Spaeth database more than 95% of the time. *See* Joseph Bafumi et al., *Practical Issues in Implementing and Understanding Bayesian Ideal Point Estimation*, 13 POL. ANALYSIS 171, 181 (2005).

[86] *See* Daniel E. Ho, Congressional Agency Control: The Impact of Statutory Partisan Requirements on Regulation (May 2007), http://law.bepress.com/alea/17th/art73.

[87] *See supra* notes 46–47 and accompanying text.

ideology does not hold true, the estimates from the model may be misleading.

Another problem with agnostic models is that they cannot make use of unanimous opinions, since it is impossible to draw any inference about the relative positions of the judges from the voting alignment in a unanimous decision. This should not be a liability if unanimous opinions truly do not convey any information, for instance, if they are cases in which the law is controlling and ideology does not play a role. However, there are many courts in which voting takes place under a "norm of consensus," so that dissent is discouraged or possibly even forbidden.[88] This is certainly true of the federal courts of appeals[89] and has been shown to be true in the U.S. Supreme Court in earlier eras.[90] In such contexts, the use of an agnostic coding model requires the researcher to discard a large proportion of the data, which may include many cases in which ideology did in fact have an impact on the outcome.

Finally, results derived from agnostically coded data may be more difficult to interpret substantively, since ideology in this context only measures how often particular judges vote with each other, and not how often they support particular types of outcomes. A judge may be deemed to be "liberal" because she frequently votes with certain liberal judges, but it would be impossible to know from agnostically coded data how often she supports the rights of criminal defendants or political protestors.

To recap, Part IV.A of this Article has identified three common approaches to the coding of case outcomes for empirical analysis, and to the definition of "ideology" itself. The "narrow" approach focuses upon judges' propensities to rule in a particular direction in a certain class of cases. The "broad" approach measures propensities to rule in a "liberal" or "conservative" direction, as defined subjectively by the

[88] *See* Ruth Bader Ginsburg, *Remarks on Writing Separately*, 65 WASH. L. REV. 133, 135–36, 138, 142–43 (1990); M. Todd Henderson, *From Seriatim to Consensus and Back Again: A Theory of Dissent*, 2007 SUP. CT. REV. 283, 292–341; Joshua B. Fischman, Estimating Preferences of Appellate Judges: A Model of "Consensus Voting" (Mar. 20, 2009), http://ssrn.com/abstract= 361348.

[89] *See* Fischman, *supra* note 88, at 1.

[90] *See* Lee Epstein et al., *The Norm of Consensus on the U.S. Supreme Court*, 45 AM. J. POL. SCI. 362 (2001).

researcher, in a broad class of cases. The "agnostic" approach, finally, measures the propensities of judges to align themselves with certain other judges.

In theory, these approaches need not yield different findings. Specifically, if ideology were purely one-dimensional, the differences among them would be irrelevant. It would not matter if judicial "liberalism" were defined in terms of case outcomes or voting alignments: a judge who favors liberal outcomes would also be one who votes with other liberal judges. Nor would it matter if cases were examined by issue area or aggregated over many areas of case law. In practice, however, these approaches do yield different results, and they do so precisely because ideology is never perfectly one-dimensional. Moreover, the more that actual judicial ideology deviates from the assumption of unidimensionality, the more that the results of empirical analysis will be sensitive to the manner in which case outcomes are coded.

B. Types of Ideology Measures

1. Proxy Measures

a. Party Affiliation and Other Simple Proxy Measures

As previously discussed,[91] the ideology of a judge cannot be observed directly. What motivates judges to decide cases in certain ways is a combination of attitudes, beliefs, and experiences that cannot be measured in the same objective fashion as a physical phenomenon. Thus, to understand the impact of ideology in empirical studies, scholars have often resorted to using "proxy variables"—variables that are thought to be correlated with a judge's underlying ideology. Various studies have, for instance, examined the impact of gender,[92] race,[93] religion,[94] and prior professional background[95] on

[91] *See supra* Part II.A.

[92] For a more detailed discussion of this literature, see Christina L. Boyd et al., Untangling the Causal Effects of Sex on Judging (July 28, 2007), http://ssrn.com/abstract=1001748.

[93] *See, e.g.*, Adam B. Cox & Thomas J. Miles, *Judging the Voting Rights Act*, 108 COLUM. L. REV. 1 (2008).

judicial decisions, on the premise that such characteristics may be correlated with the judge's policy preferences across a broad range of issues.

A particularly obvious and convenient proxy for a judge's ideology is that of membership in a political party. The linkage between a judge's party affiliation and his or her voting behavior has long been established.[96] One of the earliest empirical studies to examine differences among judges by party affiliation dates back to 1959, when Glendon Schubert analyzed decisions in workmen's compensation cases from the Michigan Supreme Court and found that judges who belonged to the Democratic Party were substantially more likely to favor employee claimants in these cases.[97] Two years later, Stuart Nagel published a comprehensive study in which he examined differences in voting behavior among the nation's nearly three hundred state and federal supreme court justices.[98] He found jurists who identified themselves as Democrats to be significantly more liberal than those who identified themselves as Republicans in every issue area he examined, including criminal law, administrative law, civil liberties, tax, family law, business, and personal injury.[99]

The most popular proxy for a judge's ideology, however, has been the party of the official who appointed the judge. The enduring popularity of this measure most likely derives from a combination of two factors. First, the party affiliation of the President or other elected official responsible for appointing a particular judge is easy both to observe and to interpret. Second, the correlation between party of appointing official and judicial ideology has long been

[94] *See, e.g.*, Gregory C. Sisk et al., *Searching for the Soul of Judicial Decisionmaking: An Empirical Study of Religious Freedom Decisions*, 65 OHIO ST. L.J. 491 (2004); Gregory C. Sisk & Michael Heise, *Judges and Ideology: Public and Academic Debates About Statistical Measures*, 99 NW. U. L. REV. 743, 759–64 (2005).

[95] *See, e.g.*, 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, 1420–21, 1470–80 (1998); Sisk et al., *supra* note 94, at 608–12.

[96] *See* GLENDON A. SCHUBERT, QUANTITATIVE ANALYSIS OF JUDICIAL BEHAVIOR (1959); Sheldon Goldman, *Voting Behavior on the United States Courts of Appeals, 1961–1964*, 60 AM. POL. SCI. REV. 374 (1966); Stuart S. Nagel, *Political Party Affiliation and Judges' Decisions*, 55 AM. POL. SCI. REV. 843 (1961).

[97] *See* SCHUBERT, *supra* note 96, at 129–42.

[98] *See* Nagel, *supra* note 96, at 843.

[99] *See id.* at 845–46.

observed over a variety of courts, time periods, and issue areas: Democratic appointees are typically more liberal on a variety of issues than Republican appointees.

The appointing-party measure has been especially dominant in studies of the federal courts. As of 1999, one paper had identified forty-one empirical studies that examined differences by party of appointing president on the circuit courts, and twenty-five such studies on the district courts.[100] Although a comprehensive treatment of this literature would be far beyond the scope of this Article, it would suffice to say that party of appointment has been shown consistently to be a statistically significant predictor of votes in most types of cases in the courts of appeals, but is less consistently correlated with judicial decision-making in the district courts.[101]

Among the many studies dating back several decades that have identified a relationship between party of appointing president and judicial voting on the federal courts of appeals,[102] a recent study by Cass Sunstein, David Schkade, and Lisa Ellman[103] has perhaps attracted the most public attention.[104] The authors examined the influence of party of appointment on decision-making in the courts of

[100] *See* Daniel R. Pinello, *Linking Party to Judicial Ideology in American Courts: A Meta-Analysis*, 20 JUST. SYS. J. 219, 225–27 tbl.1, 230–31 tbl.2 (1999). Not all of the studies canvassed by Pinello employed party of appointing president as a proxy variable in a regression analysis.

[101] *See id.* at 236 tbl.3.

[102] *See, e.g.*, Frank B. Cross, *Decisionmaking in the U.S. Circuit Courts of Appeals*, 91 CAL. L. REV. 1457, 1479–81 (2003) (reviewing scholarly use of the party proxy in studies of judicial ideology over the preceding thirty years); Sheldon Goldman, *Voting Behavior on the United States Courts of Appeals Revisited*, 69 AM. POL. SCI. REV. 491, 497 n.24 (1975) (finding a significant relationship between party of appointing president and the voting behavior of appeals court judges in the areas of criminal procedure and civil rights); Donald R. Songer, *The Policy Consequences of Senate Involvement in the Selection of Judges in the United States Courts of Appeals*, 35 W. POL. Q. 107, 111 (1982) (finding a significant relationship between judicial voting patterns and the party of both the appointing President and the judge's home-state Senators).

[103] *See* Cass R. Sunstein et al., *Ideological Voting on Federal Courts of Appeals: A Preliminary Investigation*, 90 VA. L. REV. 301 (2004); *see also* CASS R. SUNSTEIN ET AL., ARE JUDGES POLITICAL?: AN EMPIRICAL ANALYSIS OF THE FEDERAL JUDICIARY (2006) (expanding upon the earlier article).

[104] *See* Sisk & Heise, *supra* note 94, at 754–55 (describing the political attention attracted by the Sunstein et al. study); *id.* at 756 (noting that "[w]hat is perhaps most remarkable about Sunstein's study . . . is that it is *not* remarkable," in that it did not "break[] new ground in either methods or conclusions" but rather built upon several decades of earlier scholarship).

appeals across fourteen issue areas and found a statistically significant relationship in eleven of them.[105] They also found that the voting behavior of individual judges was significantly correlated with the appointing party of that judge's panel colleagues in nine of these issue areas.[106] Earlier[107] and subsequent[108] studies alike have revealed similar party and panel composition effects in administrative law cases.

Party of appointment has been shown to be a much weaker proxy, however, for judicial ideology in studies of the federal district courts. One of the most careful studies of district courts, which examined federal civil rights and prisoner cases in three district courts, found that party of appointment was not a statistically significant predictor of how the judges ruled.[109] But another large-scale study of district court decisions in criminal and civil liberties cases found that party effects changed over time: party differences were insignificant from 1960 through 1968, but Democratic appointees were significantly more liberal in the period between 1969 and 1976.[110] A more recent study of criminal sentencing found that, controlling for other relevant factors, Democratic appointees tend to give shorter sentences for serious crimes.[111]

Although party of appointment and other proxy variables can be useful and significant predictors of judicial voting, several important caveats are in order, both for those who employ such variables in their own research, and for those who encounter them in the

[105] *See* Sunstein et al., *supra* note 103, at 318–28. The exceptions were criminal appeals, federalism, and takings. *Id.* at 325–27. The latter two of these categories suffered from a small sample size; with more data, party of appointment may well have risen to statistical significance. *Id.*

[106] *See id.* In abortion and capital punishment cases, a judge's own appointing party, but not the appointing party of the other two judges on a panel, was a significant predictor of how the judge would vote. *Id.* at 328.

[107] *See* Cross & Tiller, *supra* note 39, at 2168–75; Revesz, *supra* note 38, at 1719.

[108] *See* Thomas J. Miles & Cass R. Sunstein, *Do Judges Make Regulatory Policy?: An Empirical Investigation of* Chevron, 73 U. CHI. L. REV. 823 (2006).

[109] *See* Orley Ashenfelter et al., *Politics and the Judiciary: The Influence of Judicial Background on Case Outcomes*, 24 J. LEGAL STUD. 257, 276 (1995).

[110] *See* C.K. Rowland & Robert A. Carp, *A Longitudinal Study of Party Effects on Federal District Court Policy Propensities*, 24 AM. J. POL. SCI. 291, 296 (1980).

[111] *See* Max M. Schanzenbach & Emerson H. Tiller, *Reviewing the Sentencing Guidelines: Judicial Politics, Empirical Evidence, and Reform*, 75 U. CHI. L. REV. 715, 727 (2008).

scholarship of others. First, proxy variables such as party of appointment should not be misinterpreted as *causal* variables. A judge does not think, "I was appointed by a Republican President; therefore, I should take the conservative position." Having been appointed by a President of a particular party does not cause judges to possess a particular ideology. Rather, it will correlate with the ideology that judges already possess, to the extent that Republican Presidents are more likely to appoint conservative Justices, and vice versa for Democratic Presidents.

Second, it is obvious that party of appointment and other proxies can be rather crude measures of ideology. Few studies have attempted in a systematic way to estimate how well or poorly it captures ideology.[112] Its failures, however, are sometimes highly visible: Justices Stevens and Scalia were both appointed by Republican Presidents, for example, but are viewed as occupying opposite ends of the ideological spectrum on the current Supreme Court. Scholars have offered a variety of reasons why the appointing-party proxy may fall short in practice.[113] Political actors may not be motivated entirely by ideological concerns when selecting judges;[114] even if they do care primarily about ideology, they may not have perfect knowledge of the ideology of those whom they appoint; and even if they know exactly how a candidate thinks today, they cannot necessarily predict how that candidate will behave twenty or thirty years from now.[115]

A third problem with party of appointment, which has perhaps received less attention, is that its inherent inaccuracies produce results that are systematically biased toward *understating* the impact

[112] *See infra* Part IV.

[113] *See, e.g.*, FRANK B. CROSS, DECISION MAKING IN THE U.S. COURTS OF APPEALS 20 (2007) (identifying various reasons why party of appointing President is a problematic measure of judicial ideology); Lee Epstein & Gary King, *The Rules of Inference*, 69 U. CHI. L. REV. 1, 83–84, 95–96 (2002).

[114] *See, e.g.*, David S. Law, *Appointing Federal Judges: The President, the Senate, and the Prisoner's Dilemma*, 26 CARDOZO L. REV. 479, 484–85 (2005) (noting that Presidents weigh a variety of factors, such as professional competence, political patronage, personal friendship, and electoral considerations, when selecting judicial nominees).

[115] *See, e.g.*, Epstein et al., *supra* note 64, at 1519–20 (discussing the phenomenon of "ideological drift," wherein the ideological preferences of Justices can shift in sometimes unexpected ways over time).

of ideology. As a means of estimating how much the ideology of judges affects their rulings, the appointing-party proxy is best interpreted as providing only a lower bound on ideology. For example, if Democratic appointees are 20% more likely to rule in a liberal direction in certain kinds of cases than Republican appointees, then clearly the identity of the judge must have an impact in at least 20% of cases. However, since an estimate based on a party variable only measures the difference between the *average* Democratic appointee and the *average* Republican appointee, the proportion of cases in which the identity of the judge is likely to make a difference to the outcome will be much greater.

To present a highly stylized example, assume that there are only two kinds of judges: "liberals," who *always* vote in a liberal direction, and "conservatives," who *always* vote in a conservative direction. Suppose that 80% of Republican appointees are "conservative" and 80% of Democratic appointees are "liberal." Then a regression on the party of appointment will reveal that Democratic appointees choose the liberal outcome 80% of the time and Republican appointees choose the liberal outcome 20% of the time. This difference, however, clearly understates the impact of ideology, since the difference between "conservatives" and "liberals" is 100%. This tendency toward understated results is not specific to the appointing-party proxy, but rather applies to proxy variables in general: when a proxy for some underlying variable is used in a regression to predict some phenomenon, the results will typically understate the impact of the underlying variable on the phenomenon. This problem is known to statisticians as "attenuation bias."[116]

Fourth, while the dichotomous nature of party of appointment and other popular proxy variables such as race and sex makes them easy to interpret, that same simplicity also limits what they can be used to study. Most notably, they cannot capture gradations of ideology. Party of appointment can be used to identify aggregate differences between appointees of the two parties, but it cannot distinguish between moderate and extreme judges. Consider, for example, a study by Joel Waldfogel, who found that moderate district judges

[116] *E.g.*, JEFFREY M. WOOLDRIDGE, ECONOMETRIC ANALYSIS OF CROSS SECTION AND PANEL DATA 73–76 (2002).

were more likely to induce settlement than extreme district judges.[117] Such an effect could never have been identified using the appointing-party proxy, since there is no way to identify moderates using only a dichotomous measure of ideology.

Notwithstanding these caveats, there are a number of important lessons to be gleaned from regression analysis that uses party of appointment to predict judicial voting behavior. First, such analysis has refuted the formalistic notion that judicial decision-making is the result of legal reasoning untainted by the influence of ideology and other personal characteristics. If judges decide cases simply by applying legal principles in a neutral way, there is no reason why party of appointment should correlate with judicial outcomes. Yet it does. Second, analyses using party of appointment have yielded the practical knowledge that the outcomes of cases depend to a significant extent on both the characteristics of the judge and, in the case of multimember courts, the composition of the panel. Whatever doctrinal and normative legal scholars may have to say about how judges *ought* to decide cases, practicing lawyers and social scientists want to know how judges *will in fact* behave. In a world in which reliable, useful clues about future behavior can be difficult to obtain, party of appointment is a simple and readily available piece of information with real predictive value. Third, regressions on party of appointment can reveal the nature and extent of differences between the average Republican appointee and the average Democratic appointee. Such knowledge is clearly important for understanding the impact of presidential elections on the composition of the federal judiciary, and for understanding longer-term trends in the courts.

b. Composite Proxy Measures

The perceived inadequacies of party of appointment and other simple proxy measures have led some scholars to seek superior alternatives in the form of what we might call composite proxy measures.[118] Proxy measures of this type strive for greater accuracy

[117] *See* Joel Waldfogel, *The Selection Hypothesis and the Relationship Between Trial and Plaintiff Victory*, 103 J. POL. ECON. 229 (1995).

[118] *See, e.g.*, Epstein & King, *supra* note 113, at 83–84, 95–96 (questioning the use of

by combining more than one type of information, or information from multiple sources, about the ideology of a given judge.

Until recently, a composite measure of ideology known in the literature as the "Segal-Cover scores" was also the dominant proxy for ideology in studies of the United States Supreme Court.[119] This measure was derived from newspaper editorials in four major newspapers on each Supreme Court nominee prior to confirmation. The Segal-Cover score measures the rate at which these editorials ascribe "liberal," "moderate," or "conservative" positions to the nominees.[120] Like party-of-appointment and common space scores,[121] these measures have been shown to have a statistically significant correlation with the voting behavior of the Justices. Moreover, they are also predetermined—they are fixed at the moment of confirmation—so that the direction of causality is clear: the measures of ideology are predicting differences in voting behavior among the Justices. Nevertheless, these measures should not be viewed as causal in the traditional sense; newspaper editorials preceding a Justice's confirmation do not have a direct impact on that Justice's votes. As with other proxy variables, a Justice's Segal-Cover score is *correlated* with the Justice's ideology, which is the true causal variable.

More recently, scholars have expressed enthusiasm for an ideology measure known as a "common space score" that combines a more precise estimate of presidential ideology with information about the preferences of home-state Senators to reflect the influence of the latter in the federal judicial appointments process.[122] The common

appointing party as a measure of ideology on the ground that it ignores crucial information that can help predict judicial behavior, and arguing in favor of the use of the "common space scores" devised by Giles et al., cited above in note 61).

[119] *See* Segal & Cover, *supra* note 61.

[120] *See id.* at 559–61.

[121] *See infra* notes 122–28 and accompanying text (discussing the common space scores).

[122] *See* Giles et al., *supra* note 61, at 627–31 (devising and explaining the common space scores); *see also* CROSS, *supra* note 113, at 19 (characterizing the Giles et al. common space scores as the "best currently available measure for circuit court judicial ideology" on account of the fact that they combine measures of both presidential and senatorial preference); Epstein & King, *supra* note 113, at 83–84, 95–96 (arguing in favor of using common space scores); Lee Epstein et al., *The Judicial Common Space*, 23 J.L. ECON. & ORG. 303 (2007) (rescaling the Martin-Quinn scores to correspond to the same space as the common space scores); *infra* Part IV.A.4 (discussing, and questioning, whether common space scores perform better in practice

space scores rely on sophisticated measures of senatorial and presidential ideology initially developed by political scientists Keith Poole and Howard Rosenthal.[123] The Poole-Rosenthal scores locate Senators in a two-dimensional space on the basis of the positions that they take in roll call votes, but only the first of the two dimensions is salient for most purposes.[124] The ideology scores of Presidents are then estimated along this same dimension based on the public positions that they take on bills before Congress.[125] Using the Poole-Rosenthal scores, Professors Giles, Hettinger, and Peppers proceed to assign ideology scores to federal judges as follows. First, if a judge has a single home-state Senator of the same party as the appointing President, the judge's common space score is taken to be equal to that of the Senator. Second, if both home-state senators are of the same party as the President, then the judge's common space score is the average of the two Senators' scores. Third, if both home-state Senators are of the opposite party as the President, then the judge's common space score is equal to that of the President.[126]

This approach thus incorporates substantially more information about the political circumstances surrounding the selection of a judge than just the party affiliation of the appointing President. It accounts for the ideology of a judge's home-state Senators, which is a statistically significant predictor of a judge's voting behavior owing to the practice of senatorial courtesy.[127] The scores also allow for the fact that Presidents and Senators of the same party vary in their ideological intensity, such that a Carter appointee is ordinarily presumed to be more liberal than a Clinton appointee, and so forth. The common space scores can therefore be characterized as a

than the simpler alternative of party of appointing president).

[123] *See* POOLE & ROSENTHAL, *supra* note 47, at 12–30.

[124] *See* POOLE & ROSENTHAL, *supra* note 47, at 55–62; Giles et al., *supra* note 61, at 631 (noting that the second of the two dimemsions in the Poole-Rosenthal scores rise to importance only "in a few historical eras").

[125] *See* Nolan M. McCarty & Keith T. Poole, *Veto Power and Legislation: An Empirical Analysis of Executive and Legislative Bargaining from 1961 to 1986*, 11 J.L. ECON. & ORG. 282, 297–305 (1995).

[126] *See* Giles et al., *supra* note 61, at 631.

[127] *See id.*; Donald R. Songer & Martha Humphries Ginn, *Assessing the Impact of Presidential and Home State Influences on Judicial Decisionmaking in the United States Courts of Appeals*, 55 POL. RES. Q. 299, 302 (2002).

"composite proxy" for judicial ideology, in the sense that they combine multiple sources of information—in this case, information about the ideology of both the nominating President and the home-state Senators who enjoy de facto veto power—to arrive at a measure of a judge's ideological position.

As with simpler proxy measures such as party of appointment, a major advantage of composite proxy variables is their ease of use: they are readily available, fixed at the time of appointment, and often correlate at least roughly with ideology. However, they are also harder to use than simple proxy measures: the results that they produce are less intuitive and more difficult to interpret. In a regression on party affiliation, for example, the results provide an estimate of the difference between the average Democrat and the average Republican. But the estimated effect of common space scores does not have a similarly straightforward interpretation. Although common space scores can be used quite easily to demonstrate that ideology is a statistically significant predictor of judicial voting, it is more challenging to interpret the magnitude of the effect of ideology when measured in this way.

In theory, composite proxy measures such as the common space scores should compensate for their complexity by offering greater precision than simpler proxy measures. Surprisingly little is known, however, about their actual performance across different settings. It is unclear, for example, whether and under what conditions common space scores actually outperform party of appointment as a predictor of judicial voting.[128] At the same time, composite proxy measures are subject to many of the same inherent limitations as simpler proxy measures. For example, the fact that common space scores are fixed at the time of confirmation means that they cannot account for changes in judicial ideology over time. As Martin, Quinn, and several coauthors have argued, many Supreme Court Justices "drift" ideologically during their time in office.[129] No static proxy measure—

[128] Part IV of this Article compares the accuracy of the party proxy and common space scores using a data set of asylum adjudications. We find that common space scores provide a small improvement in predictive power, but arguably not enough of a difference to justify their use in light of the difficulties in interpretation.

[129] *See* Martin & Quinn, *supra* note 32; Epstein et al., *supra* note 64, at 1502–19.

no matter how carefully devised—can adequately explain votes by these drifting Justices throughout their entire careers.

Other problems with proxy measures can arise when they are used in the positive political theory literature to study strategic interactions among judges, or between the judiciary and other institutions.[130] For a variety of reasons, such studies often require the researcher to identify the median (or most liberal, or most conservative) judge on a court or panel. An easy way to do so is to choose the judge who has the median (or most liberal, or most conservative) proxy measure of ideology. Unfortunately, such an approach runs the risk of exacerbating measurement error: on top of the measurement error inherent in any proxy measure of a given judge's ideology, there arises the further risk that one has simply selected the wrong judge. On a three-judge panel, for example, the judge with the median common space score may be the *most likely* to be the median judge, but will not necessarily be so. Nevertheless, several studies have sought to identify the median (most liberal or most conservative) judge on a court simply by singling out the judge with the median (most liberal or most conservative) proxy measure score.[131]

2. Behavioral Measures

Proxy methods have achieved such dominance in empirical studies of judicial behavior that it is easy to forget that alternatives exist. In fact, there is a large literature that has employed a variety of

[130] *See* David S. Law, *Introduction: Positive Political Theory and the Law*, 15 J. CONTEMP. LEGAL ISSUES 1, 1–2 (2006) (describing the "positive political theory of law" approach to the study of how lawmakers behave).

[131] *See, e.g.*, CROSS, *supra* note 113, at 166, 172–75 (using common space scores to identify the median, most liberal, and most conservative judges on federal appeals court panels); Jennifer L. Peresie, *Female Judges Matter: Gender and Collegial Decisionmaking in the Federal Courts*, 114 YALE L.J. 1759, 1772–73 nn.52–53 (2005) (using a rescaled version of the common space scores to identify the most liberal colleague on a given panel); *see also* VIRGINIA HETTINGER ET AL., JUDGING ON A COLLEGIAL COURT 50–51 (2006) (using absolute value of difference between common space scores as a proxy for ideological distance between judges on a panel); *cf.* Lee Epstein & Carol Mershon, *Measuring Political Preferences*, 40 AM. J. POL. SCI. 261, 275–77 (1996) (describing the failure of the Segal-Cover scores to identify the median Justice for most terms); Andrew D. Martin et al., *The Median Justice on the U.S. Supreme Court*, 83 N.C. L. REV. 1275, 1295–96 (2005) (specifically noting the probability that the Justice with the median score is in fact the median).

empirical methods to estimate the ideology of judges on an individual basis from their actual voting behavior. Whereas proxy measures of ideology rely upon generalizations from a judge's characteristics, behavioral measures of ideology demand observation of each judge's actual behavior.[132] Notwithstanding how common such approaches have been in practice, however, the scholarly literature has failed to depict them as offering an integrated and coherent alternative to proxy-based measurement approaches. One reason is that the relevant literature spans a wide range of topics and disciplines over a long period of time. Behavioral assessment of judicial ideology is an intellectually diverse tradition that ranges from much of the political science literature on the Supreme Court, to pioneering work from the earliest years of quantitative judicial politics, criminal sentencing, and empirical law and economics. Consequently, it is all too easy for scholars with a substantive interest in one area to remain unaware of what methods have been attempted or developed in another area.

Another reason that behavioral assessment remains poorly understood is that the relevant methods vary greatly in their sophistication and ambition. As a result, it is easy to lose sight of the fact that they belong to the same family of measurement approaches. At one extreme lies simple vote-counting: one can arrive at a rough assessment of a specific judge's ideological preferences in a particular context—for example, criminal law or asylum law—simply by calculating the proportion of the time that the judge in question has actually voted in favor of the defendant or asylum seeker. At the opposite extreme lies dynamic ideal point estimation of the type performed by Professors Martin and Quinn, who employ modern computing power and Bayesian simulation techniques to arrive at estimates of the Justices' "ideal points" that vary over time.[133] Their approach exploits the information conveyed by voting alignments in actual Supreme Court cases to determine where the Justices stand relative to one another. Owing to its methodological sophistication, their model can estimate the extent to which a given Justice's ideological preferences have drifted over time, for example, or the

[132] Credit belongs to Lewis Kornhauser for suggesting the "behavioral measure" terminology.

[133] *See* Martin & Quinn, *supra* note 32.

probability that a particular Justice was the ideological median of the Court at a given point in time.

Regardless of how it is implemented, the fundamental advantages and limitations of behavioral assessment remain largely the same. Its primary advantage is that it avoids the measurement error associated with proxy variables. Depending upon the proxy that is used, this measurement error may be severe: party of appointing president, for instance, is only roughly correlated with judicial behavior. Moreover, it can be difficult to know how severe this measurement error is, and how much it biases the results of a study. By contrast, even though one cannot perfectly ascertain a judge's ideology from his or her behavior, at least the extent of the measurement error can be inferred using standard statistical techniques. The primary disadvantage of behavioral assessment is that it can only be used on data that contains a relatively large number of observations per judge. If a judge only appears once in a data set, it is fruitless to try to draw any direct inference about his or her preferences. In such situations, it is necessary to resort to proxy methods, which remain capable of revealing aggregate differences in voting by party or other characteristics.

Scholars have, on occasion, rejected the behavioral-assessment approach for reasons that are not justified. There are, in particular, two misconceptions about the approach that have found their way into the judicial behavior literature. One common misconception is that proxy measures must be used because ideology itself is unobservable.[134] The whole point of statistical inference as a scientific discipline, however, is to draw reliable inferences about unobservable (or "latent") variables from whatever data can in fact be observed. There exists no "direct measure" of the extent to which judges appointed by a Democratic President are liberal, any more than there exists a "direct measure" of each judge's true ideology. The problem is not that judges' ideologies are latent variables; they are no more or less observable than any other variable to be estimated

[134] *See, e.g.*, Richard L. Revesz, *Ideology, Collegiality, and the D.C. Circuit: A Reply to Chief Judge Harry I. Edwards*, 85 VA. L. REV. 805, 823–24 (1999) (arguing that because there exists no "direct measure of the ideology held by each of the D.C. Circuit judges at the time of their appointment . . . there is no alternative but to rely on some proxy for ideology").

in a statistical model. The problem is, rather, that the methods involved in estimating the ideology of individual judges from their voting behavior can be more complicated than simply running regressions on party affiliation.

The second common misconception about the behavioral-assessment approach is that it is circular to rely upon estimates of a judge's ideology that are derived from his or her actual votes. Professors Segal and Cover have stated this objection plainly: "One cannot demonstrate that attitudes affect votes when the attitudes are operationalized from those same votes."[135] Certainly, it would be circular to estimate judges' ideologies from a data set of decisions, then use those ideology estimates to "predict" the same votes. But the purpose of most empirical research on courts is not merely to predict actual voting; it is instead to test hypotheses about why judges and litigants behave the way they do.[136] Behavioral-assessment models are often well-suited to testing hypotheses of this variety.[137]

At the same time that the limitations of behavioral assessment have at times been exaggerated, its advantages over a proxy-based approach have not always been fully appreciated. A behavioral approach often succeeds where traditional proxy methods are simply incapable of producing useful measurements of judicial ideology or testing hypotheses about judicial behavior. Although a complete survey of the literature is not possible here, a few examples should illustrate the point. One simple application that requires estimation of judicial ideology on an individualized basis from actual behavior is the study of inter-judge disparities. Ever since the Legal Realists argued that judicial decisions depend heavily upon the personal

[135] Segal & Cover, *supra* note 61, at 558; *see also* Martin et al., *supra* note 131, at 1295–96.

[136] *See* Thomas W. Merrill, *The Making of the Second Rehnquist Court: A Preliminary Analysis*, 47 ST. LOUIS U. L.J. 569, 592 (2003) (observing that, contrary to what "the traditional attitudinalists have sometimes suggested," "there is nothing 'tautological' about seeking to ascertain a Justice's preferences by examining his or her voting record and opinions as a Justice," and characterizing "each decision of the Court" as a source of "information" from which generalizations about the "revealed beliefs and attitudes of the Justices" can be drawn).

[137] If a hypothesis can be formulated as a relationship between latent variables, those variables can be included in the behavioral-assessment model, and that hypothesis can then be tested at the same time as the ideology estimates are generated using the same model. For examples of such work, see notes 138–51 and accompanying text below.

preferences of the judge, researchers have long sought to measure the differences between judges. One of the earliest known empirical studies of judges, published in 1933, uses behavioral assessment to compare the rates at which six New Jersey trial judges imposed sentences of imprisonment in criminal cases assigned to them randomly from the same pool.[138] Although the statistical analysis was rudimentary, the differences reported were large and statistically significant: the imprisonment rates among the judges varied from 34% to 58%, in a data set of over 7,000 cases.[139]

Methods that examine the behavior of individual judges are also necessary to measure changes in inter-judge disparity over time. For example, Anderson, Kling, and Stith estimate the impact of the Federal Sentencing Guidelines on inter-judge sentencing disparities within district courts.[140] Exploiting the fact that cases are randomly assigned within districts, they treat the judge assigned to each case as a "random effect" that influences the sentence imposed on the defendant.[141] Using a sophisticated model that can precisely estimate the distribution of these judge effects, they can estimate the extent to which inter-judge sentencing disparities decreased following the imposition of the guidelines.[142] Such an estimate would be impossible to derive using a proxy-based approach to the measurement of judicial ideology. Although it might be possible to estimate a lower bound on sentencing disparity by estimating the difference between the average Democratic judge and the average Republican judge, it would be impossible to draw any conclusions about changes in inter-judge disparity over time. Suppose that one were to observe that the estimated impact of party affiliation on sentence length decreased following adoption of the Sentencing Guidelines. There are two competing ways of explaining such a finding: it is possible either that inter-judge disparity decreased without regard to party affiliation, or that inter-judge disparity stayed the same while the correlation

[138] Frederick J. Gaudet et al., *Individual Differences in the Sentencing Tendencies of Judges*, 23 J. CRIM. L. & CRIMINOLOGY 811 (1933).

[139] *Id.* at 816.

[140] *See, e.g.*, James M. Anderson et al., *Measuring Interjudge Sentencing Disparity: Before and After the Federal Sentencing Guidelines*, 42 J.L. & ECON. 271 (1999).

[141] *Id.* at 290–92.

[142] *Id.* at 279–87.

between party affiliation and sentencing behavior weakened.[143] Without some way of measuring the ideology of individual judges, it would be impossible to identify the correct explanation.

Measurement of judicial ideology on an individualized basis may also be indispensable to successful empirical work when ideology has a subtle, nonlinear, or otherwise complex impact on outcomes. Professor Waldfogel's study[144] of settlement bargaining in the U.S. District Court for the Southern District of New York illustrates the point. His paper tests one empirical implication of the Priest-Klein hypothesis[145]—namely, that cases assigned to moderate judges are more likely to be litigated, while cases assigned to more extreme judges, be they highly liberal or highly conservative, are more likely to be settled.[146] By comparing settlement and decision rates among litigated cases for each judge, he identifies a clear but nonlinear relationship between ideology and settlement rates that confirms the hypothesis.[147] Although this finding is interesting in its own right, what matters from a methodological perspective is that Professor Waldfogel could not have tested this hypothesis at all had he relied upon a dichotomous proxy measure of ideology that merely distinguished liberals and conservatives, such as party of appointing president. His research question required him to distinguish not simply between liberals and conservatives, but also between liberals and conservatives, on the one hand, and moderates, on the other. Moreover, although it is conceivable that he could have executed his study using a continuous proxy measure capable of capturing the difference between moderate and extreme judges, such as the common space scores, the extra measurement error inherent in proxy measures might well have obscured the results that emerged clearly from examination of each judge's behavior.

[143] Of course, any combination of these two explanations is also possible.

[144] Waldfogel, *supra* note 117.

[145] *See* George L. Priest & Benjamin Klein, *The Selection of Disputes for Litigation*, 13 J. LEGAL STUD. 1 (1984).

[146] Waldfogel, *supra* note 117, at 299–30. In this context, the ideological spectrum can be conceived as a continuum from pro-plaintiff to pro-defendant attitudes; an "extreme" judge would exhibit a strong tendency to favor plaintiffs or defendants in certain categories of civil cases, while a "moderate" judge would exhibit neither tendency.

[147] *See id.* at 242–45.

Individualized behavioral assessment of judicial ideology is also likely to prove helpful when the dimensionality of judicial preferences is at issue. For example, in a recent study, Professor Farnsworth examined votes by Supreme Court Justices in two different types of criminal appeals, constitutional and statutory.[148] Among the nonunanimous decisions that he studied, he found a striking pattern: each of the Justices was almost exactly as likely to vote in a pro-defendant fashion in the statutory cases as in the constitutional cases, notwithstanding the very different substantive questions and legal materials at stake.[149] Overall, their voting patterns across the two types of criminal appeals were 97% correlated with one another.[150] This near-perfect correlation enables us to draw two important inferences about how the Justices vote in criminal cases. First, they appear to be guided by the same set of policy views or attitudes in constitutional cases as in statutory cases. Second, their decision-making appears to be driven more by these policy views or attitudes than by the very significant differences in formal character and content that exist between constitutional and statutory rules.

Notably, Professor Farnsworth's study did not require him to perform any regressions or complex statistical analyses; nor did he have to resort to any technically sophisticated measures of ideology.[151] What he did have to do, however, was to assess each Justice's voting tendencies in both types of criminal appeals. By doing so, he was able to discern that the ideology of the Justices can be captured by a single dimension in both types of criminal appeals: a Justice's ideological propensity to vote in favor of the criminal defendant is not neutralized or outweighed by any conflicting set of preferences having to do with the type of legal claim involved. Yet there was no way of determining that the Justices hold unidimensional ideological preferences in all criminal cases without

[148] Ward Farnsworth, *Signatures of Ideology: The Case of the Supreme Court's Criminal Docket*, 104 MICH. L. REV. 67, 68 (2005).

[149] *Id.* at 70.

[150] *Id.*

[151] *See id.*; *see also* Ward Farnsworth, *The Role of Law in Close Cases: Some Evidence from the Federal Courts of Appeals*, 86 B.U. L. REV. 1083 (2006) (showing similar correlations for circuit court judges in non-unanimous cases).

assessing the pro-defendant voting tendencies of each Justice in each category of cases.

3. Transplanted Measures

A third approach to measuring judicial ideology for purposes of empirical research is to transplant behavioral measures derived from one context into new applications. Because these measures are derived from judges' voting behavior, they are likely to be more accurate than proxy variables derived purely from political or demographic variables. It is also possible to construct such measures that may vary over time, reflecting the ideological "drift" that may occur over long judicial careers.[152] In addition, such ideology scores can be easily used in regression analysis, without any need for sophisticated modeling on the part of the researcher.

Although these "transplanted measures" seem to offer the best of both worlds—combining much of the precision of behavioral assessment with the simplicity of proxy variables—they are also easy to misuse. The most fundamental concern is that of circularity: such measures should not be used to explain voting behavior in the same data set that was used to derive them.[153] If the ideology measures are derived from the same cases being examined, it would be impossible to conclude that ideology had a causal effect on voting behavior. As Professors Epstein and Mershon put it: "To measure the political preferences of legislators by their votes at year 1 and, then, to use those very votes to explain their behavior at year 1 is to argue that legislators vote the way they do because they vote the way they do."[154]

A second trap for the unwary is that all behavioral measures of judicial ideology are derived under specific assumptions that may be inapplicable or inappropriate to the context in which a later researcher wishes to apply them. For example, if the ideology scores are derived from a model that assumes sincere voting, it would be inappropriate to use these scores to test a hypothesis that involves

[152] *See* Martin & Quinn, *supra* note 32, at 135; Epstein et al., *supra* note 64, at 1502–04.

[153] *See infra* notes 175–77 and accompanying text.

[154] Epstein & Mershon, *supra* note 131, at 262.

strategic voting. Similarly, if estimates of judicial ideology are derived from a model that employs agnostic coding of case outcomes, one must be careful when using these measures to predict outcomes in particular areas of law. Suppose, for instance, that an agnostically derived measure of ideology identifies a particular judge as liberal. Such a measure cannot necessarily be used to test the hypothesis that the judge voted in a liberal fashion in any specific area of law, such as securities fraud.[155] The judge's ideology may be multidimensional, such that she is liberal in all areas except securities fraud.[156] Alternatively, the researcher's definition of what it means to be "liberal" in this context could itself be problematic: a "liberal" judge could conceivably be animated in such cases by solicitude for the rights of defendants, or instead by hostility toward wealthy executives who enrich themselves at the expense of employee pension funds and small investors, or by some amalgam of conflicting impulses. A researcher who nevertheless uses an agnostically derived measure of ideology to predict the judge's votes in securities fraud cases thus runs the risk of concluding incorrectly that ideology is a weak predictor when, in fact, the measure may not accurately reflect the judge's ideology in this area, or the researcher may have an understanding of what it means to be liberal in a particular context that is inconsistent with how liberal judges actually think.

Transplanted measures vary greatly in their conceptual and technical complexity. The simplest variety of such measures—namely, a judge's past voting record—may not, at first glance, seem like a transplanted measure at all. To use a judge's past voting record on a particular issue to measure the current ideological preferences of that same judge on the same issue, however, is in fact to engage in a form of transplantation: one is using data from one context, that of the past, to shed light upon the judge's state of mind in a different context, that of the present. Research supports the intuition that a

[155] *See supra* Part III.A.3.

[156] *Cf. supra* text accompanying notes 53–59 (discussing the problems involved in measuring the ideology of Canadian Supreme Court Justice Claire L'Heureux-Dubé, whose agnostically derived ideology score identifies her as conservative, despite her very liberal voting record in labor and aboriginal cases).

judge's past voting record should be an excellent predictor of his or her future voting behavior: one study has found, for example, that past voting behavior is a better predictor of future voting by Supreme Court Justices in many areas of case law than the Segal-Cover scores.[157] A judge's voting history is by no means, however, a foolproof measure of his or her ideological tendencies in today's cases. It will be accurate only if a judge's ideology is stable over time, and if the questions that the judge faces now are no more likely to elicit conservative or liberal behavior than the questions he or she faced in the past.[158]

Among the most commonly used transplanted measures of ideology are the Martin-Quinn scores.[159] These scores are estimated from a one-dimensional model of voting using data from all Supreme Court cases decided since 1953.[160] Professors Martin and Quinn employ an agnostic coding methodology, meaning that the ideology measures are derived from voting alignments, rather than outcomes. Consequently, all unanimous opinions in the Supreme Court are dropped from the analysis.[161] A noteworthy strength of the Martin-Quinn approach is that it allows for, and seeks to capture, ideological movement or "drift" over time on the part of the Justices. Nevertheless, their approach still has its limitations. Perhaps most significantly, it assumes sincere voting on the part of the Justices, notwithstanding the evidence that Justices can and do vote strategically.[162]

Another criticism that has been leveled against their approach is that it is unable to distinguish between changes in the ideology of the Justices and changes in the composition of the Supreme Court's agenda over time.[163] Consider two scenarios: one in which all of the

[157] *See* Epstein & Mershon, *supra* note 131, at 275.

[158] *See* SEGAL & SPAETH, *supra* note 15, at 320–21; Symposium, The Supreme Court and the Attitudinal Model, 4 L. & CTS. 3, 3–5 (1994).

[159] *See generally* Martin & Quinn, *supra* note 32.

[160] *Id.* at 138.

[161] *See id.* at 145.

[162] *See, e.g.*, LEE EPSTEIN & JACK KNIGHT, THE CHOICES JUSTICES MAKE (1998); MALTZMAN ET AL., *supra* note 16; WALTER F. MURPHY, ELEMENTS OF JUDICIAL STRATEGY (1964).

[163] *See* Michael A. Bailey, *Comparable Preference Estimates Across Time and Institutions for the Court, Congress, and Presidency*, 51 AM. J. POL. SCI. 433, 436 (2007).

Justices drift to the right, and another in which the Justices do not change, but the Court's docket consists increasingly of cases that are inherently easier for legal and factual reasons to decide in a conservative direction. Both scenarios could conceivably produce exactly the same set of voting alignments. Thus, an approach that relies exclusively upon data on voting alignments cannot distinguish the two scenarios.

Professors Martin and Quinn do not resolve this particular problem; instead, they assume that the ideological characteristics of the agenda, as opposed to those of the Justices, are static.[164] Yet most scholars would likely agree that the Court's agenda does indeed change over time: one might plausibly suspect, for example, that the Rehnquist Court heard more cases that push doctrine in a conservative direction than the Burger Court. The inability of the Martin-Quinn approach to model changes in the Court's agenda is therefore likely to distort comparisons between Justices from different time periods and may, in particular, fail to capture the full extent to which the Court has shifted to the right over time.

To solve the problem, Professor Bailey has devised an alternative approach that bears some similarities to the Martin-Quinn model but also attempts to incorporate agenda change. Bailey's solution is to exploit "bridge observations"—namely, cases that were decided by the Court at different times, but which posed the same legal issue.[165] In effect, he attempts to control for differences in the Court's agenda over time by identifying cases that posed comparable questions at different points in the Court's history: for example, under his approach, the voting behavior of the Justices who participated in *Roe v. Wade*[166] is directly compared with that of the different set of Justices who participated in *Planned Parenthood v. Casey*.[167] By comparing how Justices at different times vote in response to the same question or issue, his approach seeks to obtain crucial information about the extent to which the agenda, as opposed to the

[164] The authors do not make this assumption explicitly, but it follows from the fact that the case parameters are drawn from a static distribution, whereas the Justices' ideal points are generated through a dynamic process. *See* Martin & Quinn, *supra* note 32, at 139–40.

[165] *See* Bailey, *supra* note 163, at 438–40.

[166] 410 U.S. 113 (1973).

[167] 505 U.S. 833 (1992).

Justices, has evolved over time. Using the same approach, Bailey further attempts to estimate ideology scores that are comparable across institutions. He does so by identifying bridge observations that span the Supreme Court, Congress, and the White House, in the form of issues and cases that were the subject of disagreement among the three branches.[168] Thus, at least in theory, the Bailey scores hold the promise of more accurate comparisons of ideology over time, and even across branches of government, than the Martin-Quinn scores. It remains the case, however, that neither measure accounts for the possibility of strategic voting.

The Martin-Quinn and Bailey measures have proved to be immensely useful to researchers on account of their accuracy and ease of use. Both sets of scores seem largely consistent with commonly held conceptions of where various Justices have stood, but closer examination does reveal a number of anomalies. For example, the Martin-Quinn scores show that the Supreme Court had its most conservative median during the 1972 Term[169]—the same Term in which it legalized abortion by a 7–2 vote in *Roe v. Wade*,[170] and only one Term after it had struck down the death penalty in *Furman v. Georgia*.[171] The Martin-Quinn scores also indicate, however, that the Court median was far more liberal in 2004 than in 1972.[172] There may be few propositions on which the entire American constitutional law professoriate can agree, but we suspect that if one were to ask each of them in which Term the Court was more conservative, they would overwhelmingly agree that it was in 2004. By contrast, the Bailey scores show that the Court gradually became more conservative since 1973, consistent with conventional wisdom.[173] Yet Bailey's approach yields its own anomalies as well: for instance, his estimates indicate that neither Justice Scalia nor Justice Thomas was to the right of Justice Rehnquist in any year that they served together on the Court.[174] If there is any other proposition on which the

[168] *See* Bailey, *supra* note 163, at 438–40.
[169] *See id.* at 436.
[170] 410 U.S. 113 (1973).
[171] 408 U.S. 238 (1972).
[172] *See* Bailey, *supra* note 163, at 436.
[173] *Id.* at 444.
[174] *See* MICHAEL A. BAILEY, IDEAL POINT DATA (2007), http://www9.georgetown.edu/

constitutional law professoriate would agree, it would surely be that Justices Scalia and Thomas were in fact the most conservative members of the Rehnquist Court.

Like any transplanted measure, the Martin-Quinn and Bailey scores must be used with caution in regression analysis. In many applications, there is a potential for circularity against which scholars have warned in the past: some, if not all, of the cases being studied will also have been used to derive the scores themselves.[175] Martin and Quinn have argued that circularity of this type will not be a meaningful problem as long as the cases being analyzed constitute only a small fraction of the data used to calculate the scores in the first place.[176] Neither the extent nor the substantive impact of circularity in any given study, however, will be easy for readers to assess. An obvious but demanding solution in such situations is simply to estimate a set of scores using the Martin-Quinn or Bailey methodology but omitting the cases under study.[177]

A second issue with sophisticated measures of the Martin-Quinn or Bailey variety is that they are easily misinterpreted. Some studies, for example, have sought to measure the ideological "distance" between Justices simply by calculating the difference between their Martin-Quinn scores.[178] Unlike simpler measures of ideology that are commonly used in the empirical literature, however, the Martin-Quinn and Bailey ideology scores are reported on numerical scales that have no natural interpretation. For example, the "distance" between Justices Alito and Thomas as measured by the difference in their Martin-Quinn scores as of the 2006 Term was 2.845, while the

faculty/baileyma/Data.htm (follow "Ideal Points" hyperlink) (online appendix to Bailey, *supra* note 163).

[175] *See supra* notes 153–54 and accompanying text.

[176] *See* Andrew D. Martin & Kevin M. Quinn, Can Ideal Point Estimates Be Used as Explanatory Variables? 2–3 (Oct. 3, 2005), http://mqscores.wustl.edu/media/resnote.pdf.

[177] *See, e.g.*, Lee Epstein et al., *The Supreme Court During Crisis: How War Affects Only Non-War Cases*, 80 N.Y.U. L. REV. 1, 55–56 n.241, 90 n.363 (2005); Matthew Sag et al., *Ideology and Exceptionalism in Intellectual Property—An Empirical Study*, CAL. L. REV. (forthcoming 2009).

[178] *See, e.g.*, Lee Epstein & Tonja Jacobi, *Super Medians*, 61 STAN. L. REV. 37, 74–81 (2008) (using the difference between Justices' Martin-Quinn scores as a measure of ideological disagreement); Nancy Staudt et al., *On the Role of Ideological Homogeneity in Generating Consequential Constitutional Decisions*, 10 U. PA. J. CONST. L. 361, 377 (2008) (using the standard deviation of Martin-Quinn scores to measure ideological homogeneity).

"distance" between Justices Alito and Souter was 2.871. Although these figures are almost identical, most observers would likely agree that Justice Alito is ideologically closer to Justice Thomas than to Justice Souter.[179] The ideological "distances" that one obtains by subtracting one score from another are misleading for the simple reason that the scale employed by the Martin-Quinn scores is not linear, but rather reports larger differences in scores at both extremes.

A third problem is that the Martin-Quinn and Bailey scores cannot be used to test or challenge the very assumptions upon which they are based. Most notably, both the Martin-Quinn and Bailey scores are estimated on the assumption that the Justices sincerely vote their ideological preferences. It is therefore inappropriate to use them to test hypotheses about strategic behavior or models that assume strategic behavior. This problem can arise, for example, when scholars seek to test models of interaction between the judiciary and other branches of government. Various studies have tackled the question of whether the Supreme Court is constrained by Congress in statutory or constitutional cases, such that it behaves more cautiously than it would otherwise do in anticipation of how Congress might react.[180] Suppose that the Justices do in fact feel constrained by the potential reactions of a conservative Congress and respond strategically by tacking to the right. In this situation, the influence of a conservative Congress might manifest itself in the Martin-Quinn or Bailey scores in the form of an ideological "drift" to the right on the part of the Justices. If so, however, then it would be inappropriate to use either set of scores to control for judicial ideology in a multivariate regression that seeks to isolate the effect of Congress on the behavior of the Court.[181] The strategic response of the Justices to the conservatism of Congress may already be reflected to some

[179] In the 2006 term, Justice Alito agreed with Justice Thomas 71% of the time in non-unanimous cases, but agreed with Justice Souter only 44% of the time. These figures were computed from the Spaeth database, cited above in note 69, with analu = 0 and dec_type = 1 or 7.

[180] *See supra* note 65 and accompanying text.

[181] *See, e.g.*, Harvey & Friedman, *supra* note 65, at 548 (using Bailey scores to measure ideology in a study of strategic voting); Sala & Spriggs, *supra* note 27, at 203 (using Martin-Quinn scores in a study of strategic voting). Both studies also report alternative estimates based on static ideology scores.

degree in their ideology scores. To include a control variable that (unbeknownst to the researcher) already reflects strategic reaction to congressional restraint is to run the risk of concluding falsely that the Court does not respond to congressional restraint.

A preferable approach is to construct a purpose-built ideology measure that is tailored to the study at hand and does not embody an answer to the very question that the researcher intends to test. Thus, for example, to test whether the Supreme Court is more constrained in matters of statutory interpretation when it faces a hostile Congress, Professor Segal constructs a novel measure of ideology that is specifically suited for his task: he derives a measure of the Justices' ideology solely from their voting behavior in constitutional cases.[182] If one is willing to assume that the Court is constrained by Congress only in statutory cases and not in constitutional ones, this measure offers an appropriate baseline for measuring the ideological preferences of the Justices in the absence of congressional constraint.

Similarly, Professors Epstein, Ho, King, and Segal construct a custom measure of judicial ideology to examine the impact of an ongoing war on the Supreme Court's decision-making in civil liberties cases.[183] It would be inappropriate to use the Martin-Quinn scores to control for the impact of judicial ideology because the scores are estimated in large part from the very civil liberties cases that the authors are studying. Consequently, the authors employ two alternative measures of judicial ideology—the Segal-Cover Scores, which are static and do not depend upon the actual voting behavior of the Justices, and a set of scores calculated using the Martin-Quinn algorithm, but from a subset of cases that includes no civil liberties decisions.

IV. COMPARISON OF MEASUREMENT METHODS

Although there are a wide variety of methods available for measuring ideology, surprisingly little research has been done to

[182] *See, e.g.*, Jeffrey A. Segal, *Separation-of-Powers Games in the Positive Theory of Congress and the Courts*, 91 AM. POL. SCI. REV. 28 (1997); *see also* Bergara et al., *supra* note 65, at 247–80 (employing Segal's ideology measure in an alternative econometric model to criticize his conclusions).

[183] *See* Epstein et al., *supra* note 177, at 90 n.363.

assess their strengths and weaknesses.[184] In this Part, we compare a few of the measurement approaches that have proven most popular in the empirical literature on the Supreme Court and the federal courts of appeals. Our goal is not to compare all available measures, but rather to demonstrate the difficulties involved in choosing and employing a measure of ideology that is appropriate to a particular context. It is hoped that our demonstrations will prove useful both to researchers with a methodological interest in the design of quantitative studies, and to readers who wish to understand the limitations of such studies.

A. How Different Methods Perform When Applied to the Federal Courts of Appeals

Our application of different measurement methods to the courts of appeals will make use of two separate contributions by each of the authors of this Article. We will compare the performance of three different approaches to the measurement of judicial ideology on the courts of appeals using Law's data set of asylum adjudications in the Ninth Circuit.[185] First, we examine decision-making in the asylum cases using party of appointing president as a proxy for ideology. Next, we analyze the same data using another proxy measure, the "judicial common space" scores.[186] The third measurement approach that we consider is a behavioral approach, Fischman's "consensus voting" model, which estimates the ideology of each judge from his or her voting behavior while also accounting for the influence of collegial interaction in multi-member courts.[187]

[184] *E.g.*, Sisk & Heise, *supra* note 94, at 787–90; *see also* Epstein & Mershon, *supra* note 131, at 262.

[185] *See* David S. Law, *Strategic Judicial Lawmaking: Ideology, Publication, and Asylum Law in the Ninth Circuit*, 73 U. CIN. L. REV. 817 (2005) [hereinafter Law, *Strategic Judicial Lawmaking*]; David S. Law, *Judicial Ideology and the Decision to Publish: Voting and Publication Patterns in Ninth Circuit Asylum Cases*, 89 JUDICATURE 212 (2006).

[186] *See supra* notes 122–28 and accompanying text. Common space scores for circuit court judges are available at http://epstein.law.northwestern.edu/research/JCS.html (last visited Apr. 14, 2009). We thank Lee Epstein for making this data available.

[187] *See* Fischman, *supra* note 88. Although a full treatment of the underlying model is beyond the scope of this Article, it assumes that judges vary along an underlying ideological dimension that ranges from completely pro-asylum to completely anti-asylum. It further assumes that judges incur disutility from dissenting, and it modifies its prediction of how each

1. Description and Initial Exploration of the Data

The data set contains all 1,892 asylum appeals decided by ordinary three-judge panels of the Ninth Circuit from 1992 through 2001.[188] Judges were deemed to have voted in a "pro-asylum" direction if they joined an opinion that favored any kind of relief for the asylum petitioner; otherwise, their votes were deemed to be "anti-asylum." As coded in this manner, 18% of the votes cast were pro-asylum, while 82% were anti-asylum. Forty-eight percent of all judicial votes in the data set were cast by Democratic appointees and 52% by Republican appointees. The overwhelming majority of the cases were decided unanimously (95%) and without a published opinion (92%).[189]

Table 1 illustrates the relationship between party of appointment and judicial voting. It reports the rates at which judges cast pro-asylum votes, broken down by the appointing party of both the voting judge and his or her colleagues. It should be immediately evident that party of appointment is strongly associated with voting behavior in asylum cases: the average Democratic appointee is 13% more likely to vote in favor of relief than the average Republican appointee. The party of appointment of a judge's colleagues also has a large impact on how he or she will vote. Republican appointees have a 6% pro-asylum voting rate when sitting on all-Republican panels, but that rate rises to 20% when they sit with two Democratic appointees. Similarly, Democratic appointees exhibit a 15% pro-asylum voting rate when no other judge on the panel is a Democratic appointee, which increases to 25% when both of the judge's colleagues are fellow Democrats.

judge will vote to reflect this "cost of dissent." The model yields estimates of each judge's ideology as well as the cost of dissent, which can then be used to predict the probability of a pro-asylum vote for each judge on a panel. *See id.*

[188] Law, *Strategic Judicial Lawmaking*, *supra* note 185, at 832.

[189] *See id.* at 817, 855.

TABLE 1: PERCENTAGE OF PRO-ASYLUM VOTES CAST

	No panel colleagues are Democratic appointees	One panel colleague is a Democratic appointee	Both panel colleagues are Democratic appointees	Average across all scenarios
Republican appointee	6%	12%	20%	12%
Democratic appointee	15%	25%	35%	25%
Average across all judges	10%	18%	27%	18%

The fact that party of appointment is correlated with how judges vote in certain types of cases does not necessarily mean, however, that it is an accurate measure of judicial ideology. To be sure, it is indispensable for certain purposes. If, for example, our goal were to evaluate the impact that changes in control of the White House have on the ideological direction of the federal courts,[190] then it would be both obvious and appropriate to study the relationship between party of appointment and judicial voting. But suppose instead that our goal is to understand the extent to which the ideological composition of a panel affects how judges vote. In that case, we would be interested not in the impact of appointing party per se, but rather in the impact of the actual ideology of the judges, for which appointing party is merely an imperfect proxy. As explained above in Part III.B.1, a proxy measure such as party of appointment that is imperfectly correlated with the underlying variable of interest—in this case, judicial ideology—will tend to understate the true impact of the underlying variable.[191] A minority of Republican appointees may be relatively liberal; likewise, some fraction of Democratic appointees may in fact be somewhat conservative. Both a liberal Republican appointee who votes liberally and a conservative Democratic appointee who votes conservatively are behaving ideologically. If one uses party of appointment to measure their ideology, however, one will mistakenly conclude that neither judge is voting his or her

[190] *See* DEBORAH J. BARROW ET AL., THE FEDERAL JUDICIARY AND INSTITUTIONAL CHANGE 12 (1996) (reporting that "[t]he combination of new positions and swelling numbers of vacancies, owing especially to retirements," has "enabled modern presidents to change anywhere from 35 to 60 percent of the membership on the lower federal courts during their stay in office").

[191] *See supra* note 116 and accompanying text.

ideological preferences. The result will be a systematic failure to capture the full impact of ideology.

2. Regression Analysis and Goodness-of-Fit Comparison

We evaluated the performance of the three measurement approaches at issue—namely, party of appointing president, judicial common space scores, and Fischman's consensus voting model—by performing three regressions. In all three regressions, the dependent variable was the direction (either pro-asylum or anti-asylum) of a particular judge's vote in a given case. The regressions on appointing party and common space scores were estimated using a random-effects probit model, which is well-suited for data with dichotomous outcomes and also accounts for the fact that the three judges on a panel may be influenced by unobserved factors specific to each case.[192] Such factors might include, for example, facts about the asylum petitioner or claim that were known to the judges but are unavailable to the researcher.

In the first two regressions, the independent variables were the ideology measure of the judge casting the vote in question and a measure of the ideology of the other two judges on the panel. The ideology scores of the judge's colleagues were included in order to

[192] *See* WOOLDRIDGE, *supra* note 116, at 485–86 (2002). The random effects assumption is justified by the fact that cases are randomly assigned to panels. Many papers in the empirical literature on the courts of appeals have employed regression models, typically logit or probit, that make the implausible assumption that the three votes in each case are independent. *See, e.g.*, Revesz, *supra* note 38, at 1767; Cross & Tiller, *supra* note 39, at 2169–70; Sunstein et al., *supra* note 103, at 316 n.40. Although such models will still estimate the regression coefficients correctly, the standard errors may be incorrect unless clustered. *See* WOOLDRIDGE, *supra* note 116, at 482; Sag et al., *supra* note 177. However, for the purpose of estimating probabilities of particular outcomes, rather than merely estimating regression coefficients, the random effects model is more appropriate. As a robustness check, we also estimated the regressions in this paper using a plain probit model, and the results were very similar.

The random-effects probit regression on the common space scores takes the form:

$$\text{Pr(Pro-asylum vote by judge } i \text{ in case } n) = \Phi[b_1 CS_i + b_2(CS_j + CS_k) + c_n],$$

where CS_i denotes the common space score of judge i, CS_j and CS_k denote the common space scores of the other two judges on the panel, c_n is a normally distributed random effect, b_1 and b_2 are coefficients to be estimated, and Φ denotes the cumulative density function of the normal distribution. The regression on the party proxy uses a party indicator in place of the common space scores.

account for the collegial nature of panel voting. Circuit court panels are commonly thought to operate under a "norm of consensus" such that, all other things being equal, judges are generally reluctant to dissent, especially in low-profile cases.[193] Thus, the ideology of a judge's colleagues may influence his or her vote. In the first regression, the ideology of the voting judge was coded as a "1" if he or she was appointed by a Democratic President and "0" otherwise. The measure of the ideology of the other judges on the panel was simply the number of Democratic appointees among the remaining judges on the panel, ranging from 0 to 2. For the second regression, we substituted the "common space" measures of judicial ideology for party of appointing president: the independent variables were the voting judge's own common space score, and the sum of the common space scores of the judge's two colleagues. The third regression, in which we estimated Fischman's consensus voting model, was considerably more complex, as it called for the estimation of almost eighty parameters, including the asylum voting proclivities of most judges in the Ninth Circuit.

Table 2 presents the results of the first two regressions. The first column reports the results of the regression in which party of appointment was used as a proxy for the ideology of both the voting judge and the other judges on the panel. The second column reports the results of the regression in which the judicial common space scores were used in lieu of appointing party. In both regressions, both the ideology measures for the voting judge and for the other judges on the panel are statistically significant at the $p \leq .01$ level. The estimated effects are also in the expected direction. Being a Democratic appointee, and having panel colleagues who are Democratic appointees, increase the likelihood of a pro-asylum vote. Likewise, the common space scores correspond as expected to the voting behavior of the judges: liberal judges, as measured by the common space scores, are more likely to favor asylum relief than conservative ones. The estimates from both regressions suggest that 45% of the ideological component of a judge's vote is determined by

[193] *See* POSNER, *supra* note 15, at 32–34; *see also* SUNSTEIN ET AL., *supra* note 103, at 64–71 (providing several alternative explanations for panel composition effects).

the judge's own ideology and 55% is determined by the ideology of the other two judges.

The common space scores do a slightly better job of explaining how the judges voted, as measured by the log-likelihood and pseudo-R^2 goodness-of-fit statistics for each regression, but the difference is very small. Moreover, the pseudo-R^2 for both models is relatively low, which suggests that ideology—at least, as measured by these proxy variables—explains only a small proportion of the variation in voting. By comparison, the pseudo-R^2 for Fischman's consensus voting model is dramatically greater, which is to say that it is much better at explaining the voting behavior of the judges.[194] This is only to be expected, however, given that it employs a vastly greater number of explanatory variables.[195]

TABLE 2: EFFECT OF IDEOLOGY ON THE LIKELIHOOD OF A PRO-ASYLUM VOTE

	Party of appointment	Common space scores
Democratic appointee	1.63**	
Number of colleagues who are Democratic appointees	1.00**	
Common space score		-2.49**

[194] Fischman's model, which estimates the ideology of the judges individually, yields a pseudo-R^2 of 0.324 compared to 0.061 and 0.067 for the party and common space models, respectively. However, these statistics are not directly comparable, as the third model contains seventy-nine independent variables, as opposed to only three independent variables in each of the proxy-based models.

[195] Fischman's consensus voting model includes a parameter for each judge–whether active, senior, or sitting by designation–who participated in at least ten asylum cases in the Ninth Circuit between 1992 and 2001. There were sixty-five parameters of this type in total. In addition, the model includes two parameters that capture the average ideology of Democratic and Republican appointees who participated in fewer than ten cases; one parameter representing the ideological variance within these groups; one parameter that reflects the "cost of dissent"; and a parameter to represent the variance of the case-specific random effect. Thus, the model employs a total of seventy-nine independent variables, whereas each of the proxy-based models employs only three independent variables.

	Party of appointment	Common space scores
Sum of colleagues' common space scores		-1.55**
Constant	-4.74**	-2.98**
Goodness of fit statistics:		
Log likelihood	-1398.28	-1389.48
Pseudo-R^2	0.061	0.067

** denotes statistical significance at the $p \leq .01$ level

Although log-likelihood and pseudo-R^2 statistics are widely used measures of goodness-of-fit, they are not easy to interpret in practical terms. A more intuitive way to assess the extent to which the three types of ideology measures fit the data is to speak in terms of predicted probabilities. Each type of ideology measure can be used to predict the probability that a particular vote will be pro-asylum. Accordingly, we calculated these predicted probabilities for each vote in the data using each type of ideology measure. We then compared the average predicted probability of a pro-asylum vote in all instances when a judge actually cast a pro-asylum vote, against the average predicted probability of a pro-asylum vote in all instances when a judge actually cast an anti-asylum vote. The difference between these two averages is a measure of goodness-of-fit known as "lambda."[196] The larger the value of lambda for a particular measure of ideology, the greater the portion of the variation in voting behavior that can be explained by that measure.

For each vote in the data that was actually cast in favor of asylum, the party of appointment measure yielded, on average, a 21.8% predicted probability of a pro-asylum vote. By contrast, for each vote

[196] *See* J.S. Cramer, *Predictive Performance of the Binary Logit Model in Unbalanced Samples*, 48 STATISTICIAN 85, 88 (1999). The "lambda" statistic is appropriate for assessing the goodness-of-fit of models with dichotomous outcomes and is especially helpful when one outcome—in this case, anti-asylum votes—are much more common than the other. *See id*. Another commonly used measure, the percentage of votes correctly predicted, can be misleading in such situations. *Id.* at 91–92.

in the data that was actually cast against asylum, the party of appointment measure yielded, on average, a 16.4% predicted probability of a pro-asylum vote. The lambda goodness-of-fit statistic for the party-of-appointment measure was therefore 5.4%. For the common space scores, the corresponding lambda statistic was a somewhat better 6.3%. The ideology measures based upon Fischman's consensus voting model, however, proved far superior at distinguishing between pro-asylum votes and anti-asylum votes: pro-asylum votes had a 35.7% average predicted probability of being pro-asylum, while the anti-asylum votes had only a 14.2% average predicted probability of being pro-asylum, with a resulting lambda of 21.5%.

3. The Extent to Which Proxy Measures Understate the Impact of Ideology

It is not surprising that Fischman's implementation of the behavioral-assessment approach does a much better job of explaining the votes in the data, given that it is much more complex than either of the proxy models.[197] Our goodness-of-fit comparisons do highlight, however, a bona fide strength of behavioral measures: unlike proxy-based approaches, an approach that turns upon examination of each judge's behavior does not tend to understate the impact of ideology on judicial behavior. Proxy measures such as appointing party and common space scores are inherently imprecise and cannot fully capture the ideological voting patterns that actually exist in the data. The lambda statistic calculated above in Part IV.A.2 is merely one way of describing the extent to which they fall short in this regard.

Another way to illustrate the extent to which the proxy measures fail to capture the full impact of ideology on voting is to compare, under each approach, the predicted difference in judicial behavior between a liberal voting scenario and a conservative voting scenario. In other words: what is the difference under each approach between the likelihood that a *liberal* judge paired with like-minded colleagues will cast a pro-asylum vote, and the likelihood that a *conservative*

[197] *See supra* note 195 (discussing the number of parameters in each of the models).

judge paired with like-minded colleagues will cast a pro-asylum vote?

In the case of the appointing party proxy, the difference is easy to see. Per Table 1, a Republican appointee sitting with two Republican appointees supports asylum relief only 6% of the time, while a Democratic appointee sitting with two Democratic appointees supports asylum relief 35% of the time. Thus, there is a 29% gap in the pro-asylum voting rate between a Republican judge on an all-Republican panel and a Democratic judge on an all-Democratic panel. This gap represents the impact of ideology, as measured by party of appointment: 29% of voting in asylum cases hinges upon the appointing party of the judge and his or her colleagues. The simplicity of this analysis illustrates an important respect in which party of appointment shines as a measure of ideology: it produces results that require little effort to interpret. Moreover, it does so without any need for regression analysis or calculation of predicted probabilities and, if our goodness-of-fit tests are to be believed, at little cost in accuracy over a more complex proxy measure such as the common space scores.

To obtain comparable predictions from the common space and behavioral measures, we must again calculate predicted probabilities. Specifically, for each vote in the data, we use each of the two measures to calculate the predicted probability that the judge in question would vote in favor of asylum. The predicted probabilities for each judge may then be ranked, from the lowest (reflecting the behavior of a conservative judge paired with conservative colleagues) to the highest (reflecting the behavior of a liberal judge paired with liberal colleagues).

The results of this analysis, applied to each of the three methods for measuring ideology, are shown in Table 3, with probabilities arranged by percentile. When we calculate these probabilities using a model that employs the common space scores, we find that the judge at the 5th percentile, or conservative end, of this spectrum is only 6% likely to cast a pro-asylum vote, whereas the judge at the 95th percentile, the liberal end of the spectrum, has a 35% chance of casting a pro-asylum vote.[198] This range in pro-asylum voting rates is almost identical to the range observed when we simply compare all-Republican panels with all-Democratic panels.[199] The behavioral-assessment model, however, reveals a much larger role for ideology than does either the common space score regression or a simple appointing-party analysis. According to this model, the difference between the 5th percentile and the 95th percentile is the difference between a 1% likelihood of a pro-asylum vote and a 57% likelihood of a pro-asylum vote. In other words, judicial voting is being driven by ideology over half of the time.

TABLE 3: PREDICTED PROBABILITY OF A PRO-ASYLUM VOTE, ACCOUNTING FOR THE INFLUENCE OF OTHER PANEL MEMBERS

Probability of pro-asylum vote	Appointing party	Common space score	Behavioral assessment
5th percentile	6%	5%	1%
Median	15%	16%	11%
95th percentile	35%	35%	57%

4. The Mixed Performance of the Common Space Scores

Although there has been much debate about the relative merits of common space scores as opposed to party of appointment as measures of judicial ideology,[200] Table 3 suggests that the

[198] We compare the 5th and 95th percentile, rather than the minimum and maximum, in order to reduce sensitivity to outliers.

[199] Yet another approach would be to use the regression coefficients from Table 2 to calculate the range in voting rates using the party variable, but the results would still be unchanged. The 5th percentile judge in the data would be a Republican appointee on an all-Republican panel, and would have a 6% predicted pro-asylum voting rate. The 95th percentile judge would be a Democratic appointee on an all-Democratic panel and would have a 35% predicted pro-asylum voting rate. These predictions correspond exactly with the estimates derived from Table 2. According to the party proxy, the judge with the median voting probability would be a Democratic appointee sitting with two Republican colleagues.

[200] *Compare* Epstein & King, *supra* note 113, at 83–84, 95–96 (favoring common space scores), *and* Lee Epstein & Gary King, *A Reply*, 69 U. CHI. L. REV. 191, 203 n.27 (2002)

performance differences between these two proxy measures are quite small in practice. Common space scores provide a better fit to the asylum voting data than party of appointment, but the difference is slight. Both measures yield almost identical estimates of the impact of ideology on how judges vote. The much larger discrepancy is between the two proxy methods, on the one hand, and the behavioral-assessment approach, on the other. It is the measurement error inherent in the proxy measures that explains both the vastly inferior fit of the proxy-based models and the fact that such models substantially underestimate the impact of ideology.

Figure 1 illustrates the magnitude of the measurement problems associated with the common space scores. For each judge, we used Fischman's consensus voting model to estimate the probability that he or she would prefer to vote in favor of asylum in a random case.[201] Figure 1 plots each judge's estimated probability of preferring the pro-asylum outcome against the judge's common space score.[202] The overall correlation between the estimated probabilities and the common space scores is 0.56. Although this correlation is statistically significant, the scatterplot highlights how the common space scores do a poor job of capturing the ideology of many judges whose estimated voting behavior departs dramatically from what their common space scores would suggest: some judges with similar or even identical common space scores nevertheless have very different estimated probabilities of voting in favor of asylum.

To some extent, these disparities reflect inaccuracies in the behavioral estimates as well as flaws in the common space scores. We believe that the crux of the problem lies, however, in the common space scores. The scatterplot merely confirms problems in the

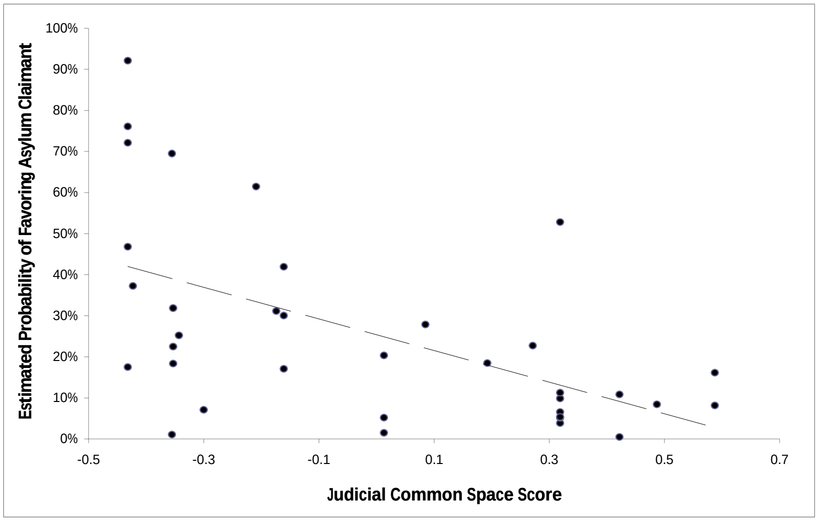
(same), *with* Frank Cross et al., *Above the Rules: A Response to Epstein and King*, 69 U. CHI. L. REV. 191, 203 n.27 (2002) (defending use of party of appointing president), *and* Richard L. Revesz, *A Defense of Empirical Legal Scholarship*, 69 U. CHI. L. REV. 169, 180–83 (2002) (same).

[201] The probability that a judge *prefers* the pro-asylum outcome is the rate at which the judge would vote in favor of asylum if she were to vote sincerely. Because our model accounts for the fact that judges do not always vote sincerely, but are instead influenced by their panel colleagues, this probability will differ from a judge's actual pro-asylum voting rate. Panel composition effects are discussed above at notes 38–44 and accompanying text.

[202] Because a judge's ideology cannot be estimated precisely from a small number of votes, Figure 1 includes only judges who cast at least fifty votes in the data.

common space scores that are obvious without any resort to estimated voting probabilities. For example, Judge Noonan, a Reagan appointee, has a very conservative common space score, yet he is in fact one of the most liberal judges on the Ninth Circuit in the area of asylum, judged either by his raw voting record–which "place[s] him among a handful of liberal Carter appointees"[203]–or by the 61% rate of favoring asylum claimants that the model predicts for him.[204] Conversely, Judge Farris is a Carter appointee with a very liberal common space score, yet he has in fact compiled an extremely conservative voting record in asylum cases,[205] and the model predicts that he would favor asylum claimants only 2% of the time.

FIGURE 1: ESTIMATED PROBABILITY OF A JUDGE FAVORING THE ASYLUM CLAIMANT, COMPARED TO HIS OR HER COMMON SPACE SCORES[206]



[203] Law, *Strategic Judicial Lawmaking*, *supra* note 185, at 851–52, 852 fig.8.

[204] *See id.* at 851–52 & 852 fig.8.

[205] *See id.* at 852 fig.8.

[206] Figure 1 includes only judges with at least fifty votes in the data. The predicted probability of a pro-asylum vote is derived from the voting data using Fischman's consensus voting model. The downward-sloping line is the best-fit line; it is the line that provides the best possible fit to the data, in the sense that it minimizes the sum of squared errors.

One might attempt to explain the poor performance of the common space scores in this context on the ground that asylum cases are somehow atypical of the kinds of cases that divide judges ideologically. On this view, Judge Noonan's scores might do a better job of capturing his ideological leanings in, say, employment discrimination cases than in asylum cases. This is simply another way of saying that judicial ideology may be multidimensional: judges who may be conservative in other contexts may not be conservative in the context of asylum, and vice versa.[207]

There is little evidence, however, that common space scores do a better job of explaining how judges behave in other types of cases. Their use has been justified primarily on theoretical grounds, rather than on the basis of any practical demonstration that they explain judicial behavior well in any specific area of law.[208] There have been few other empirical tests of common space scores, and the results have been decidedly mixed. In a study of religious freedom cases decided by the lower federal courts, the common space scores barely outperformed party of appointing president in the context of Establishment Clause cases.[209] These results were reversed, moreover, when it came to a substantial category of Free Exercise Clause cases, with party of appointing president slightly outperforming the common space scores.[210] Similarly, in a recent study of sex discrimination cases decided by the federal courts of appeals, common space scores provided a slightly worse fit than party of appointing president, although it appears unlikely that the difference was statistically significant.[211]

[207] *See supra* note 156 and accompanying text (noting that an agnostically coded measure of judicial ideology may not accurately capture ideology in every area of law if the content of a judge's ideological preferences varies from one area of law to another). As discussed above in Parts III.A.1 and III.B.2, behavioral assessment avoids this problem entirely, if the data are limited to a narrow area of law.

[208] *See* Epstein & King, *supra* note 113, at 95–96.

[209] *See* Sisk & Heise, *supra* note 94, at 789 & n.266–67.

[210] *See id.* (noting that party of appointing president performed slightly better in the context of Free Exercise Clause litigation in educational settings).

[211] *See* Boyd et al., *supra* note 92, at 44 tbl.3, 45 tbl.4 (reporting log-likelihood levels that indicate party of appointing president provided slightly better fit than common space scores when employed in a logistic regression model of judicial voting in published sex discrimination cases decided by the federal courts of appeals); E-mail from Christina L. Boyd, Ph.D. Candidate, Washington University in St. Louis, to the authors (Apr. 20, 2009, 12:21 EST) (on

The sparse and inconsistent track record of the common space scores illustrates an unfortunate fact: too little is known about the actual performance of various measures of ideology across different areas of law. To the extent that certain measures may perform well in certain applications, further research will be necessary to identify those applications.

5. A Summary of the Relevant Tradeoffs

In sum, no measurement approach is ideal in all respects. Yet the strengths and weaknesses of different approaches may not always be obvious. In our own tests, the oldest and humblest of measures—namely, party of appointing president—proved surprisingly robust against a more sophisticated competitor. Contrary to claims that common space scores constitute a superior measure of judicial ideology,[212] we found that such scores performed little better at explaining the data than party of appointment. The two proxy measures also yield practically identical estimates of the impact of ideology on judicial voting. At the same time, measurements and estimates that rely upon party of appointment have the added advantage of being easy to interpret. Behavioral assessment, in turn, produces ideology measurements that perform much better at explaining judicial voting than either party of appointment or common space scores. Compared to the other two approaches, it also does a better job of capturing the full impact of ideology on judicial behavior.

The behavioral-assessment approach, however, is often substantially more difficult to implement. Nor will its advantages always be relevant as a practical matter: the additional statistical power that it offers will not always make the difference between accepting or rejecting a hypothesis. For purposes of testing certain basic hypotheses—for instance, that the ideology of each and every judge on the panel has a statistically significant impact on how any given member of the panel will vote—all three approaches are likely

file with the authors) (confirming that the “party variable employed in the regressions refers to party of appointing president).

[212] *See, e.g.*, Epstein & King, *supra* note 113, at 95–96.

to yield results of overwhelming statistical significance. In such cases, simpler may indeed be better.

B. How Different Methods Perform When Applied to the Supreme Court

Empirical research on the Supreme Court presents unique challenges due to the relatively small number of Justices and the unique nature of its docket. The composition of the Court can remain stable for long periods, and yet individual Justices may undergo ideological drift over their careers.[213] The Supreme Court hears a disproportionately large share of high-profile and "difficult" cases, but scholars have thus far had little success in quantifying either the legal character of these cases or the substantive criteria for granting certiorari. Nevertheless, the Court has long been the subject of intensive empirical study, and scholars have devised a variety of ways to measure the ideology of its members.

The simplest measure is, once again, party of appointing president: Democratic appointees are presumed to be more liberal than Republican appointees. Although party of appointment was employed in some of the earliest empirical studies of the Supreme Court, it is now seldom used, presumably due to the availability of superior measures. The Segal-Cover scores, another commonly used measure, are derived from an analysis of newspaper editorials written about each Justice in the prelude to his or her appointment.[214] Once the dominant measure of Supreme Court ideology, they have ceded that position to the Martin-Quinn scores, which use cutting-edge techniques to estimate the "ideal points" of the Justices from their voting behavior in actual cases.[215]

We set out to compare the explanatory power of these measures using decisions drawn from five issue areas: criminal procedure, economic activity, the First Amendment, civil liberties other than those found in the First Amendment, and judicial power. Data on the

[213] *See supra* note 129 and accompanying text.

[214] *See* Segal & Cover, *supra* note 61, at 560; *see also supra* notes 119–21 and accompanying text.

[215] *See* Martin & Quinn, *supra* note 32; *see also supra* notes 159–64 and accompanying text.

relevant cases was obtained from the Supreme Court database compiled by Harold Spaeth, which includes every Supreme Court case decided on the merits from 1953 through 2006.[216] The ideology measures that we tested included not only party of appointment, the Segal-Cover scores, and the Martin-Quinn scores, but also a behavioral-assessment measure of our own devising. For each of the five areas of law at issue, we estimated the ideology of each Justice on the basis of his or her voting record in that particular area.[217] In the area of criminal procedure, an abundance of data enabled us to estimate for each Justice not only an ideology score, but also a linear time trend parameter that provides some indication of the extent and direction of the Justice's ideological movement over time.[218]

We then evaluated the performance of the four measurement approaches at issue by performing a series of regressions. In each regression, the dependent variable was the ideological direction of a Justice's vote in a case, as coded in the Spaeth database.[219] In the party of appointment, Segal-Cover, and Martin-Quinn regressions,[220]

[216] An updated version of the database that includes cases decided since 2006 is available online. *See* Supreme Court Data, *supra* note 69.

[217] This model is specified below in note 222.

[218] The modified model that we employed for the criminal procedure cases is specified in note 222 below.

[219] The coding of this variable and others like it in the Spaeth database has increasingly become the subject of scholarly criticism. *See* Edelman & Chen, *supra* note 63, at 306–07 (criticizing the manner in which legal issues are coded in the Spaeth database); Shapiro, *supra* note 63, at 501 (arguing that scholars who wish to rely upon the variables relating to substantive law in the Spaeth database face two problems: "(1) the impossibility of knowing how many (and which) legal issues arise in a particular case and (2) the difficulty of using the Database to study the way different areas of law interact or affect each other"); Harvey, *supra* note 68, at 21–22; Landes & Posner, *supra* note 68, at 42 (criticizing, and seeking to correct, the Spaeth database's coding of the ideological direction of Supreme Court decisions). Although it is important to be cognizant of such criticisms, a degree of miscoding of judicial votes in the Spaeth database is probably of only limited relevance to the comparison that we perform in this Article, as there is no obvious reason to think that any miscoding will systematically favor one measurement approach over another.

[220] The regressions of judicial voting on the Martin-Quinn scores raise a problem of endogeneity or "circularity," as some scholars have called it. *See supra* notes 153–54, 175–77 and accompanying text. The problem exists because the Martin-Quinn scores are being employed in a regression to predict a subset of the votes from which they were estimated in the first place. *See supra* notes 153–54, 175–77 and accompanying text. However, this criticism would apply to *any* use of the Martin-Quinn scores as a transplanted measure, which is precisely how the Martin-Quinn scores have typically been used in the empirical literature. Moreover, Martin and Quinn have themselves argued that endogeneity of this type is not of

the independent variables were the ideology measure and also, for the Segal-Cover and Martin-Quinn scores, a squared version of that same measure.[221] Given the difficulties involved in controlling for the unique characteristics of each case before the Supreme Court, we employed a fixed-effects logit model for each of the regressions.[222] This type of regression model allows for the possibility that the Justices' votes are influenced in a common manner by issues specific to the case, yet does not require us to quantify any details of the case itself. It does so by estimating the probability that each Justice will cast a liberal vote, conditional on the total number of liberal votes. For example, if a particular case is decided by a six-to-three margin in the liberal direction, the model estimates the impact of Justice's ideology score on the probability that he or she will cast a liberal vote, in light of the fact that six liberal votes are being cast. Because this type of model cannot generate any inferences about the impact of ideology on voting when all Justices vote the same way, we included only non-unanimous decisions in our data.

practical concern in many applications. *See* Martin & Quinn, *supra* note 176, at 2–3. The endogeneity problem would appear to be most severe when the Martin-Quinn scores are used to explain criminal procedure votes, which constitute 37% of the cases in the Spaeth database.

[221] The use of a squared term is appropriate due to the possibility of differences in the nonlinear scale of each ideology measure.

[222] *See* WOOLDRIDGE, *supra* note 116, at 491–92. The fixed-effects logit model used here differs from the random-effects probit model used in Part IV.A in that it does not rely on the assumption that cases are randomly assigned. Because the Supreme Court has a discretionary docket, the characteristics of cases selected for review will presumably vary with the ideologies of the Justices at the time that review is granted.

The regression on the Martin-Quinn scores, for example, takes the form:

$$\Pr(\text{Liberal vote by Justice } i \text{ in case } n \text{ in Term } t) = \Lambda(b_1 MQ_{it} + b_2 MQ_{it}^2 + c_n),$$

where MQ_{it} represents the Martin-Quinn score for Justice i in Term t , c_n is a case-specific fixed effect, b_1 and b_2 are coefficients to be estimated, and Λ denotes the logistic function, $\Lambda(x) = 1 / (1 + e^{-x})$. The regressions on the Segal-Cover scores and party of appointment take the same form except that these ideology measures are static, and the party-of-appointment model omits the squared term.

The behavioral-assessment model for each issue area other than criminal procedure estimates a parameter x_i for each Justice from the model:

$$\Pr(\text{Liberal vote by Justice } i \text{ in case } n) = \Lambda(x_i + c_n).$$

The model used for the criminal procedure cases estimates an additional parameter, a Justice-specific linear time trend z_i:

$$\Pr(\text{Liberal vote by Justice } i \text{ in case } n \text{ in Term } t) = \Lambda(x_i + tz_i + c_n).$$

Although the measures that we are comparing here are all considered measures of judicial "ideology," it bears repeating that they are fundamentally dissimilar at a conceptual level and thus may in fact be capturing different phenomena. The Martin-Quinn scores are derived from an agnostic voting model estimated over the entire range of cases decided by the Supreme Court. Thus, the regression of, say, criminal procedure votes on Martin-Quinn scores estimates the degree to which Justices who often vote with "liberal" Justices across all cases happen to favor defendants in criminal procedure cases specifically. By contrast, the Segal-Cover regression estimates the degree to which Justices who were identified as "liberal" by newspaper editorialists before their confirmation favor criminal defendants. The party proxy explains how much more likely Democratic appointees are to support defendants than Republican appointees. The behavioral model directly estimates each Justice's propensity to support defendants. It only makes sense to compare these measures if we believe that they all capture the same underlying phenomenon—namely, a single dimension of ideological disagreement that helps to explain the voting behavior of the Justices across a broad range of cases.

Not surprisingly, each ideology measure proved a statistically significant predictor of how the Justices voted in each of the three regressions. For the sake of simplicity and brevity, we do not report the regression coefficients, which are difficult to interpret due to the complexity of the models and the varying scales of the ideology measures. Instead, Table 4 reports for each regression the pseudo-R^2 statistic, a goodness-of-fit measure that reflects how well the predictors in each regression explained the voting decisions of the Justices.

TABLE 4: GOODNESS-OF-FIT FOR DIFFERENT IDEOLOGY MEASURES ACROSS DIFFERENT ISSUE AREAS

	Pseudo-R^2 for each measure, by area of law				
	Criminal procedure	Civil rights	Economic activity	First Amendment	Judicial power
Behavorial assessment	0.61	0.50	0.22	0.53	0.20
Martin-Quinn	0.61	0.53	0.17	0.44	0.17
Segal-Cover	0.38	0.22	0.07	0.31	0.08
Appointing party	0.06	0.06	0.03	0.03	0.03
Number of non-unanimous cases	1540	702	530	456	368

In terms of explaining the voting choices of the Justices, it is clear that the Martin-Quinn scores and the behavioral measures are superior to the Segal-Cover scores, which are in turn a large improvement over the party-of-appointment proxy measure. The fit of the measures also varies substantially by issue area. On the whole, all of the ideology measures provide a better fit for criminal procedure, civil rights, and First Amendment cases than for economic and judicial power cases, which suggests that the ideological preferences of the Justices may be unidimensional across only some areas of law: the attitudes that cause Justices to vote liberally in both civil rights and criminal procedure cases, for example, may have less relevance to their behavior when it comes to economic or judicial power cases.

The goodness-of-fit statistics alone do not convey a practical sense of the extent to which the Martin-Quinn scores outperform the Segal-Cover scores. Accordingly, Table 5 translates the regression results into a metric that is much easier to interpret: the probability of a liberal vote. Due to the computational challenges involved in estimating such probabilities, we limit ourselves to reporting the predicted probability that a given Justice would be among those casting a liberal vote in a criminal or economic case that is ultimately decided by a five-to-four margin in a conservative direction. Moreover, because the Martin-Quinn scores vary over time, we must calculate the probabilities as of a specific point in time. Accordingly, our predictions are based on the Martin-Quinn scores for the 1999 Term.[223] For purposes of comparison, Table 5 also lists the actual proportion of liberal votes cast by each Justice in five-to-four conservative rulings over the period from 1994 until 2004. Given that there are only thirty-six actual cases that meet these criteria, however, these proportions are necessarily imprecise and should be interpreted with caution.[224]

TABLE 5: PREDICTED PROBABILITY OF A LIBERAL VOTE IN CRIMINAL PROCEDURE CASES, 1999 TERM, CONDITIONAL ON A 5–4 CONSERVATIVE OUTCOME

	Behavioral assessment	Martin-Quinn	Segal-Cover	Actual proportion of liberal votes, 1994–2004
Rehnquist	0.08	0.14	0.17	0.06
Stevens	0.95	0.97	0.41	0.94
O'Connor	0.22	0.23	0.59	0.06
Scalia	0.11	0.08	0.13	0.11
Kennedy	0.19	0.20	0.54	0.03
Souter	0.79	0.75	0.50	0.92
Thomas	0.08	0.08	0.30	0.06
Ginsburg	0.85	0.85	0.71	0.97
Breyer	0.74	0.70	0.63	0.86

[223] We chose the 1999 Term because it is the midpoint of the 1994–2004 period, during which the composition of the Supreme Court remained stable. Because a Justice's Segal-Cover score does not change over time, the predictions obtained from that regression are identical for all Terms in which the composition of the Court remained the same. By contrast, the Martin-Quinn scores do vary over time, but we suspect that the predictions and goodness-of-fit measures that they yield would not vary dramatically over the 1994 to 2004 period. Calculation of those statistics for each year would, in any event, be computationally burdensome and beyond the scope of this Article.

[224] Consider, for example, the fact that Justice Kennedy voted in favor of the defendant in only 3% of the criminal procedure cases that were decided by a five-to-four conservative vote during the period from 1994 to 2004. In other words, he voted in a liberal direction in only one out of the thirty-six cases that met this description. At the same time, however, he had a 12% pro-defendant voting rate in cases that were decided by a six-to-three conservative vote, and an 11% pro-defendant voting rate in cases that were decided by an even more lopsided seven-to-two conservative vote. These figures suggest that the extremely low proportion of liberal votes reported for Justice Kennedy in Table 5 may not be representative of his overall behavior.

The behavioral measure and the Martin-Quinn scores yield almost identical predictions of how likely each Justice is to vote liberally. Perhaps this is not surprising: the Martin-Quinn scores are themselves behavior-based estimates of the Justices' ideologies, albeit derived from a different model and a broader set of cases. By contrast, however, the Segal-Cover scores frequently yield very different predictions from the Martin-Quinn scores. For example, the Segal-Cover scores predict that Justice Stevens will vote liberally only 41% of the time, compared to the 97% liberal voting rate predicted by the Martin-Quinn scores; likewise, whereas Justice Thomas has a predicted 30% liberal voting rate according to the Segal-Cover scores, the Martin-Quinn scores predict that he will vote liberally only 8% of the time.

A quick glance at the actual voting records of the Justices suggests that the Martin-Quinn scores yield predictions that are more reliable than those generated from the Segal-Cover scores. If one compares the predicted probabilities from each model with the actual voting percentages reported in the last column of Table 5, it becomes apparent that the Martin-Quinn scores yield predictions that fall closer to the observed data *for every Justice*. Meanwhile, the predictions based upon the Segal-Cover scores are sometimes far off the mark. For example, Justice Stevens has an actual record of voting liberally 94% of the time in such cases, yet the Segal-Cover scores predict that he will vote liberally only 41% of the time.

Table 6 mirrors Table 5, but in the context of economic cases: it reports the predicted probability that a given Justice will cast a liberal vote in an economic case that is ultimately decided by a five-to-four margin in a conservative direction, as of the 1999 Term. Once again, the behavioral measures and the Martin-Quinn scores yield almost identical results, but the Segal-Cover scores often produce divergent predictions. For instance, the Segal-Cover scores predict a liberal voting rate of 43% for Justice Stevens, whereas the Martin-Quinn scores predict a 90% liberal voting rate. In the context of economic cases, it is difficult to evaluate the performance of the three measurement approaches against the actual voting records of the Justices: the Court decided only eight economic cases from 1994 to 2004 by a five-to-four conservative vote, which is simply too small a number to permit meaningful comparison. Given that the Segal-

Cover scores provided substantially lower goodness-of-fit,[225] however, it is reasonable to infer that the Martin-Quinn scores and the behavioral approach to measuring ideology will yield more reliable predictions.

TABLE 6: PREDICTED PROBABILITY OF A LIBERAL VOTE IN ECONOMIC CASES, 1999 TERM, CONDITIONAL ON A 5–4 CONSERVATIVE OUTCOME

	Behavioral model	Martin-Quinn	Segal-Cover
Rehnquist	0.17	0.23	0.26
Stevens	0.87	0.90	0.43
O'Connor	0.24	0.30	0.54
Scalia	0.22	0.19	0.22
Kennedy	0.35	0.28	0.51
Souter	0.58	0.62	0.48
Thomas	0.20	0.19	0.35
Ginsburg	0.73	0.71	0.63
Breyer	0.63	0.58	0.57

In all of our regressions, the Martin-Quinn scores perform almost as well as the behavioral measures that we separately estimated for each area of law. They provide comparable goodness-of-fit in every area of law that we examined and yielded similar predictions for the voting behavior of each Justice in criminal procedure and economic cases. However, the behavioral-assessment approach still offers important advantages. Most significantly, it avoids the problem of circularity that is inherent whenever the Martin-Quinn scores are used to explain voting on the Supreme Court.[226] It can also be used to test hypotheses that cannot be tested with the Martin-Quinn scores. Consider, for example, the question of whether Justice Ginsburg was more liberal than Justice Souter in economic cases as of 1999. Per Table 6, the results of the behavioral approach support the conclusion that she was indeed more liberal, and that the difference is

[225] *See supra* Table 4.

[226] *See supra* note 220.

statistically significant.[227] However, this simple hypothesis cannot be tested using the results from the Martin-Quinn regression, because any regression that uses the Martin-Quinn scores is built on the assumption that the Martin-Quinn scores have correctly specified the ideological placement of the Justices relative to one another.[228]

The Segal-Cover scores still have one advantage over both the Martin-Quinn scores and our own behavioral measures: because they are fixed at the time of appointment and do not depend at all upon the actual voting behavior of the Justices, they can be used in any kind of regression analysis without raising concerns of circularity. At the same time, they provide reasonable explanatory power across many, albeit not all, areas of law.[229] This combination of virtues may help to explain their enduring popularity, even in sophisticated applications.[230] Our results suggest, however, that the Segal-Cover scores may not be appropriate in applications that demand a high degree of measurement precision.

CONCLUSION

Not all measures of judicial ideology are created equal. This much is already widely suspected, if not known, among those who study judicial behavior empirically. Very little has been written, however, about which measures are better, to what extent, and for what purposes. This Article has sought to rectify this situation in two ways. First, we have identified the most pressing conceptual and methodological challenges involved in measuring judicial ideology. Second, we have sought to measure the measures themselves, by evaluating the relative performance of several popular approaches to measuring judicial ideology. Our findings confirm that different

[227] Per a Wald test, the difference is statistically significant at the $p = .10$ level.

[228] *See supra* note 137 and accompanying text (discussing how a behavioral-assessment approach to measuring judicial ideology can be better suited to testing certain types of hypotheses about judicial behavior); *supra* notes 155–56 and accompanying text (explaining that agnostically coded measures of judicial ideology, such as the Martin-Quinn scores, embody assumptions that may render them inappropriate for certain applications).

[229] *See supra* Table 4 (reporting meaningful goodness-of-fit statistics for the Segal-Cover scores when used to predict judicial voting in the areas of criminal procedure, civil rights, and First Amendment law).

[230] *See* Epstein et al., *supra* note 177, at 55–57, 90.

measures of ideology vary greatly in their ability to explain judicial voting, and that the choice of one measurement approach over another can significantly influence the findings that scholars reach. If empirical scholarship involving the concept of judicial ideology is to realize its scientific potential or gain greater acceptance from a wider audience, those of us who produce such scholarship must learn both to speak clearly about what is meant by "judicial ideology," and to give careful thought to the methods that are employed to measure it. Because no measurement approach is ideal, it will inevitably be necessary for scholars to make tradeoffs and to sacrifice the advantages of one approach for the virtues of another according to the project at hand. But it is certainly possible to choose among the alternatives in an informed fashion.