The Real World of Arbitrariness Review

Thomas J. Miles† *& Cass R. Sunstein*††

INTRODUCTION

In the 1960s and 1970s, the federal courts of appeals, above all the United States Court of Appeals for the District of Columbia Circuit, developed the "hard look doctrine."[1] The doctrine found its origins in judicial decisions requiring administrative agencies to demonstrate that they had taken a "hard look" at the underlying questions of policy and fact.[2] Hence agencies were required to offer detailed, even encyclopedic, explanations for their conclusions, to respond to counterarguments, to justify departures from past practices, and to give careful consideration to alternatives to the proposed course of action.[3] These were procedural requirements, to be sure, but they had significant effects, often shifting regulatory policy in identifiable directions by, for example, discouraging the construction of nuclear power plants[4] and generally leading agencies to give heightened attention to environmental protection.[5] Eventually courts went well beyond these proce-

† Assistant Professor of Law, The University of Chicago.

†† Karl N. Llewellyn Distinguished Service Professor, The Law School and Department of Political Science, The University of Chicago. We thank Eric Posner, Richard Posner, Peter Strauss, and Adrian Vermeule for helpful comments. We are also grateful to Rachael Dizard, Casey Fronk, Darius Horton, Matthew Johnson, Bryan Mulder, Brett Reynolds, Matthew Tokson, and Adam Wells for superb research assistance.

1 For an influential and well known example, see *Ethyl Corp v EPA*, 541 F2d 1, 35 (DC Cir 1976) (en banc) (stating that a court should "undertake[] a study of the record . . . even as to the evidence on technical and specialized matters").

2 See Harold Leventhal, *Environmental Decisionmaking and the Role of the Courts*, 122 U Pa L Rev 509, 511 (1974).

3 All of these requirements can be found in *Motor Vehicle Manufacturers Association v State Farm Mutual Auto Insurance Co*, 463 US 29, 43 (1983). In the same vein, see *Corrosion Proof Fittings v EPA*, 947 F2d 1201, 1212 (5th Cir 1991) (requiring the agency to give notice of its intended methodology "while the public still has an opportunity to analyze, comment, and influence the proceedings").

4 See Stephen Breyer, Vermont Yankee *and the Courts' Role in the Nuclear Energy Controversy*, 91 Harv L Rev 1833, 1838–39 (1978) ("The licensing process, including court review, would seem at least partly responsible for the long lag between plan and operation. . . . [O]ne suspects that delay in the licensing process would tend to lead a firm to decide in favor of coal [instead of building a nuclear power plant].").

5 See William F. Pederson, Jr., *Formal Records and Informal Rulemaking*, 85 Yale L J 38, 59 (1975) (arguing that the EPA may promulgate suboptimally protective regulations due to intra-agency and external opposition, and that strong judicial review is an effective check on this pattern). For an early and illustrative signal of the intended effect of judicial review in the environmental domain, see *Calvert Cliffs' Coordinating Committee, Inc v Atomic Energy Commis-*

dural requirements to take a hard look on their own, assessing the reasonableness of agency judgments of policy and fact on their merits.[6]

The goal of hard look review was to police agency decisions for genuine arbitrariness, not to allow federal judges to impose their own policy preferences on the administrative state.[7] Indeed, a central point of judicial review was to respond to the open-ended delegation of discretionary power by ensuring a firm check on agency decisions that might be "irrational or discriminatory."[8] On this view, the hard look doctrine might be seen as a second-best substitute for the original constitutional safeguards against the uncontrolled exercise of discretion. Judicial scrutiny of agency judgments of policy and fact might even serve as a method for reducing factional power over government in a way that would recall longstanding concerns about the problems posed by the exercise of authority by self-interested private groups.[9]

As it developed, however, the hard look doctrine became highly controversial.[10] Some of the controversy involved its likely effects. Would the doctrine discourage agency action altogether, and therefore freeze the status quo rather than improving agency decisions?[11] Some of the controversy involved its legal foundations. Was hard look review an illegitimate creation of the federal courts?[12] What provision of law authorized federal judges to impose these various requirements on agen-

sion, 449 F2d 1109, 1111 (DC Cir 1971) (asserting that the role of courts in judicial review is to ensure that agencies "live up to the congressional mandate" of protecting the natural environment).

6 See, for example, *Citizens to Preserve Overton Park, Inc v Volpe*, 401 US 402, 415 (1971) (maintaining that a presumption of regularity does not protect an agency from "a thorough, probing, in-depth review").

7 See id at 416 (explaining that although there must be a "searching and careful" review, the court may only determine whether the agency made a "clear error of judgment . . . and is not empowered to substitute its judgment for that of the agency").

8 *Ethyl Corp*, 541 F2d at 68 (Leventhal concurring) (arguing that judicial review ensures that agencies promulgate objectives and regulations in a reasonable manner that is consistent with their statutory powers).

9 See Federalist 10 (Madison), in *The Federalist* 56, 59–60 (Wesleyan 1961) (Jacob E. Cooke, ed).

10 See, for example, Peter L. Strauss, *Revisiting* Overton Park*: Political and Judicial Controls over Administrative Actions Affecting the Community*, 39 UCLA L Rev 1251, 1266–68 (1992) (arguing that the Court may have improperly distrusted the political process); Jerry L. Mashaw and David L. Harfst, *The Struggle for Auto Safety* 96–97 (Harvard 1990) (criticizing circuit court interpretations of the "arbitrary and capricious" and "hard look" standards for burdening agencies and for causing uncertainty as to which issues might be "significant" at trial); Stephen Breyer, *Judicial Review of Questions of Law and Policy*, 38 Admin L Rev 363, 388–90 (1986) (arguing that judges lack the technical knowledge and independent access to information necessary to make a well-informed judgment about agency decisions).

11 See Mashaw and Harfst, *Auto Safety* at 95 (cited in note 10) (describing how judicial review invalidating agency regulations creates "massive disruption" and significant delay).

12 For discussion, see Clark Byse, Vermont Yankee *and the Evolution of Administrative Procedure: A Somewhat Different View*, 91 Harv L Rev 1823, 1829 (1978) (arguing that the creation of additional procedures is an agency, not a court, determination).

cies or to give careful scrutiny to the merits?[13] Independent questions lay in the background: Would judicial biases distort the inquiry into reasonableness?[14] Might judicial judgments reflect not an assessment of irrationality or discrimination, but the judges' own policy commitments?

In its seminal decision in *Motor Vehicle Manufacturers Association v State Farm Auto Mutual Insurance Co*,[15] the Court entrenched hard look review and clarified its foundations. The Court rooted its analysis in § 706 of the Administrative Procedure Act[16] (APA), which requires courts to strike down agency action found to be "arbitrary" or "capricious."[17] According to the Court, a decision would count as arbitrary if

> the agency has relied on factors which Congress has not intended it to consider, entirely failed to consider an important aspect of the problem, offered an explanation for its decision that runs counter to the evidence before the agency, or is so implausible that it could not be ascribed to a difference in view or the product of agency expertise.[18]

These words, quoted hundreds of times in federal court decisions,[19] were widely taken to ratify both procedural and substantive components of the hard look doctrine.[20]

Many agency decisions, including those of the NLRB, are also subject to review as lacking "substantial evidence." In its 1951 decision in *Universal Camera Corp v NLRB*,[21] the Court emphasized that the substantial evidence test of the National Labor Relations Act[22] (NLRA)

---

13 This question is raised in *Vermont Yankee Nuclear Power Corp v NRDC*, 435 US 519, 548 (1978) (finding that no statute permitted the lower court to review and overturn the rulemaking procedure if that procedure met the statutory minimum).

14 See *Ethyl Corp*, 541 F2d at 67 (Bazelon concurring) (warning that aggressive judicial review will "compound the error of the panel in making legislative policy determinations alien to its true function").

15 463 US 29 (1983).

16 5 USC § 706(2)(A) (2000).

17 Id ("The reviewing court shall . . . hold unlawful and set aside agency action, findings, and conclusions found to be . . . arbitrary, capricious, an abuse of discretion, or otherwise not in accordance with law.").

18 *State Farm*, 463 US at 43.

19 In fact, these words have been quoted in no fewer than 572 federal cases as of April 23, 2007. LEXIS search, Apr 2007.

20 See Cass R. Sunstein, *Deregulation and the Hard-look Doctrine*, 1983 S Ct Rev 177, 210 (arguing that *State Farm* endorses the primary elements of the hard look doctrine by requiring substantive consideration of the facts and supporting the procedural reversal of agency decisions in order to maintain efficient regulation).

21 340 US 474 (1951).

22 29 USC § 160(e) (2000) (requiring that the NLRB's factual findings be upheld "if supported by substantial evidence").

and the APA[23] was "a response to pressures for stricter and more uniform practice," embodying a legislative "mood" in favor of increased judicial "responsibility for the reasonableness and fairness of Labor Board decisions."[24] In practice, and especially in the aftermath of *State Farm*, review under the substantial evidence standard is essentially the same as review under the arbitrary and capricious standard,[25] though it is sometimes thought that review for substantial evidence is somewhat more searching.[26]

Since *State Farm*, the Court has issued no major pronouncements about judicial review of allegedly arbitrary agency action, and the doctrine has remained essentially stable for over two decades.[27] But the controversies that preceded the decision have continued unabated.[28]

---

23 5 USC § 706(2)(E) (2000) (stipulating that certain agency actions and findings shall be set aside if "unsupported by substantial evidence").

24 *Universal Camera*, 340 US at 487, 489–90 (exploring that mood).

25 For the substantial evidence test, see *Allentown Mack Sales & Service, Inc v NLRB*, 522 US 359, 366–67 (1998) (applying the substantial evidence standard and inquiring whether a reasonable jury could have come to the same conclusion as the NLRB); *Universal Camera*, 340 US at 481 (elaborating that standard). Some provisions of the environmental statutes also call for substantial evidence review.

The claim that there is no difference between the substantial evidence test and the arbitrary and capricious standard is supported by the fact that the legislative history of the statute in *State Farm* itself suggested that agency findings must be reviewed under the substantial evidence test. Notwithstanding that fact, the Court used the arbitrary and capricious standard—a decision that would be puzzling if the substantial evidence test were more severe. See *State Farm*, 463 US at 43–44, 46 (noting that "it is relevant" that Congress requires agency actions to be supported by substantial evidence, but that the "ultimate question" before the Court was whether the agency action was arbitrary and capricious). By emphasizing the arbitrary and capricious standard, the Court seemed to suggest that the substantial evidence standard was essentially identical.

26 *State Farm* is best taken as adopting the call for searching review issued long before in *Overton Park*, 401 US at 420 (remanding to the district court for review "based on the full administrative record that was before the Secretary at the time he made his decision").

27 For a representatively minor pronouncement, at least on the general operation of arbitrariness review, see *Verizon Communications, Inc v FCC*, 535 US 467, 502 (2002) (distinguishing between an original agency interpretation of a statute, which is evaluated under *Chevron U.S.A. Inc v NRDC*, 467 US 837 (1984), and a modified agency interpretation, to which *State Farm* would be applicable). The most important ruling involving substantial evidence review may well be *Allentown Mack*, which did seem to suggest an unusually aggressive approach, but that decision has not spurred significant rethinking in the lower courts. See Stephen Breyer, et al, *Administrative Law and Regulatory Policy: Problems, Text, and Cases* 214 (Aspen 6th ed 2006).

The absence of a major ruling from the Court is itself something of a mystery. Why have we not seen large-scale developments from the Supreme Court in nearly a quarter century? The answer may well lie in the fact that arbitrariness review is typically focused on specific questions of fact and policy, which makes Supreme Court review less likely and which also makes Supreme Court rulings less likely to turn out to be broad pronouncements. We explore this point and its implications below.

28 Though the central issues involved statutory interpretation rather than arbitrariness, the Court's decision in *Massachusetts v EPA*, 127 S Ct 1438 (2007), might well be taken as a modern version of *State Farm*, also involving a "hard look." See 127 S Ct at 1463 (holding that the "EPA has offered no reasoned explanation for its refusal to decide whether greenhouse gases cause or contribute to climate change" despite the EPA's reliance on a report from the National Research

Some people object that the doctrine has unfortunate systemic effects on agency decisions.[29] Others believe that the hard look is simply too hard and that a soft look would be much better.[30] Still others fear that judicial biases play a large role in the operation of the hard look doctrine—that in finding inadequate explanations or unreasonableness on the merits, the policy preferences of judges are playing a substantial role.[31] It is perhaps revealing here that *State Farm* itself, involving a high-profile initiative by the Reagan Administration, produced, on some key issues, what seemed to be a political division within the Court on the arbitrariness question, with conservative justices siding with the Reagan Administration.[32] And other observers, most prominently Justice Breyer, have objected that there is an evident incongruity in the fact that under existing doctrine, courts often defer to agency interpretations of law, while taking a hard look at agency judgments about policy and fact.[33]

To date, only a sparse empirical literature exists on the actual operation of the hard look doctrine.[34] There is no systematic evidence on the rate of invalidation under hard look review; we do not know if the rate is 10 percent, or 20 percent, or 40 percent. Nor is there evidence

---

Council that questioned the causal link). For a valuable discussion to this effect, see Jody Freeman and Adrian Vermeule, Massachusetts v. EPA*: From Politics to Expertise*, 2007 S Ct Rev (forthcoming) ("Before *MA v. EPA*, it was unclear whether discretionary decisions not to promulgate regulations were even reviewable, let alone subject to the 'hard look' review. In this way, *MA v. EPA* could be *State Farm* for a new generation.").

29 See, for example, Mashaw and Harfst, *Auto Safety* at 149–51 (cited in note 10) (contending that hard look review led the National Highway Traffic Safety Administration to avoid rulemaking and focus on product recalls); Breyer, 38 Admin L Rev at 391–93 (cited in note 10) (arguing that judicial requirements of an exhaustive investigation of alternatives may prevent agencies with scarce resources from making even minor changes).

30 See, for example, Richard J. Pierce, Jr., *Seven Ways to Deossify Agency Rulemaking*, 47 Admin L Rev 59, 65 (1995) (arguing that "courts have transformed the simple, efficient notice and comment process into an extraordinarily lengthy, complicated, and expensive process").

31 See, for example, R. Shep Melnick, *Regulation and the Courts: The Case of the Clean Air Act* 11–12 (1983) (asserting that judges no longer trust agencies to develop balanced policies and use judicial review to push agencies to be more protective of citizens' interests).

32 Compare *State Farm*, 463 US at 51, 55–56 (White majority, including Brennan, Marshall, Blackmun, and Stevens) (invalidating as arbitrary an agency decision with respect to detachable and nondetachable seat belts), with id at 58 (Rehnquist concurring in part and dissenting in part) (arguing that the decision on these points was not arbitrary).

33 See Breyer, 38 Admin L Rev at 394 (cited in note 10) ("Courts are fully capable of rigorous review of agency determinations of law, for it is the law that they are expert in. . . . But, for reasons of 'comparative expertise,' increased judicial scrutiny seems *less* appropriate [for policy decisions].").

34 For the principal exceptions, see Joseph L. Smith and Emerson H. Tiller, *The Strategy of Judging: Evidence from Administrative Law*, 31 J Legal Stud 61, 64 n 9, 81 (2002) (presenting evidence that strategic reasons motivate judges' choice of *Chevron* or *State Farm* as the basis for their decisions); Richard L. Revesz, *Environmental Regulation, Ideology, and the D.C. Circuit*, 83 Va L Rev 1717, 1719 (1997) (finding that ideology influences judicial voting, that the effect is stronger for cases less likely to be reviewed by the Supreme Court, and that panel effects influence voting).

on the role, if any, of judicial policy preferences. Do Republican and Democratic appointees vote differently in cases involving hard look review? Are panels with a majority of Republican appointees different from panels with a majority of Democratic appointees, and if so, how different are they?

Our aim here is to begin to fill this gap. We do so through an analysis of a large data set, consisting of all published appellate rulings from 1996 to 2006 involving review of decisions of the EPA and review of NLRB decisions either for arbitrariness or for lack of substantial evidence.[35] (For convenience, we use the phrase "arbitrariness review" to capture the relevant test, which does not seem to differ significantly across the two contexts.[36]) Use of this data set has several advantages. First, both agency and judicial decisions are fairly easy to code in political terms, and hence it is possible to test competing hypotheses about the role of judicial ideology.[37] Second, EPA and NLRB decisions are extremely important in their own right, and they also provide a good "snapshot" of the world of arbitrariness review.[38] Third, there is a large data set, in essentially the same period, involving judicial review of interpretations of law by the EPA and the NLRB.[39] An examination of arbitrariness review permits instructive comparisons.[40] Fourth, the EPA is an executive agency, whose head is an at-will employee of the president, whereas the NLRB is an independent agency, whose chair and majority are determined by the incumbent president but whose members may be discharged only for cause (and therefore have, in practice, a form of tenure). Hence our data set includes two of the most prominent agencies, one of which has the "executive agency" form and the other of which has the "independent agency" form.[41]

---

35 For arbitrariness review and the NLRB, see note 25.

36 See note 25 and accompanying text.

37 This task is far more difficult for such agencies as the FCC and the SEC, where political coding can be quite contentious. We have, however, compiled a data set of all cases citing *State Farm*, offering some preliminary conclusions about validation rates; the data are available on request.

38 Of course we cannot exclude the possibility that the patterns we observed here are different from those of other agencies.

39 See Thomas J. Miles and Cass R. Sunstein, *Do Judges Make Regulatory Policy?: An Empirical Investigation of* Chevron, 73 U Chi L Rev 823, 825 (2006).

40 The data set for *Chevron* cases extends across a larger time period, but we find consistent results across time in those cases. Hence, the comparison holds.

41 While our focus is on the period between 1996 and 2006, it is entirely imaginable that similar patterns would be found in similar periods, including those preceding *State Farm*. We would not be at all surprised, for example, if in review of NLRB cases between 1956 and 1966, broadly similar patterns might be found. Prior research on how administrative agencies fare before the Supreme Court reveals that the success rates vary substantially across agencies, but that overall agencies' success rates are generally stable over time. See note 86. It would be most interesting, however, to examine directly how the patterns we discern change, if at all, over time.

The central goal of arbitrariness review is to filter out serious errors of analysis, not to encode judicial policy preferences, and we are interested above all in testing whether courts are carrying out that task. Much of the debate in modern administrative law is about this question,[42] which has yet to be tested. If Democratic and Republican appointees show significantly different rates of "liberal voting" in cases reviewing agency decisions for arbitrariness review, there is evident reason for concern. And if all-Democratic panels show dramatically different voting patterns from all-Republican panels, there is reason to believe that similarly situated litigants are not being treated similarly, in a way that has serious consequences for regulatory policy and even the rule of law.

In brief, our principal findings are as follows:

1. Political commitments significantly influence the operation of hard look review in EPA and NLRB cases. When the agency decision is liberal, the Democratic validation rate is 72 percent and the Republican validation rate is 58 percent. When the agency decision is conservative, the Democratic validation rate drops to 55 percent and the Republican validation rate rises to 72 percent. For both Republican and Democratic appointees, then, the likelihood of a vote to validate is significantly affected by whether the agency's decision is liberal or conservative.

2. In an important sense, these figures understate the role of ideology in hard look review, because panel effects are substantial (regardless of whether the agency decision is liberal or conservative). Democratic appointees show higher liberal voting rates (75 percent) when sitting with two other Democratic appointees. Republican appointees show lower liberal voting rates (55 percent) when sitting with two other Republican appointees. The resulting difference between the two sets of appointees—20 percentage points—has large consequences for the real world of administrative law.

3. For EPA and NLRB cases, taken together, the overall rate of votes to validate agency decisions challenged as arbitrary is 64 percent. Notably, the rate at which the judges vote to validate is significantly higher for Democratic appointees than for Republican appointees: 70 percent for Democratic appointees and 60 percent for Republican appointees. Strikingly, the rate of valida-

---

42 See, for example, Breyer, 38 Admin L Rev at 395 (cited in note 10) (questioning whether judicial review improves agency decisionmaking); Melnick, *Regulation and the Courts* at 343 (cited in note 31) (commenting that "judicial discretion" leads some circuit courts to be especially favorable to environmental groups and others to be especially favorable to business interests).

> tion is essentially the same in arbitrariness review as in cases dealing with statutory interpretation[43]—a finding that casts doubt on Justice Breyer's suggestion that courts might be giving greater scrutiny to agency judgments of fact than to agency judgments of law.[44]

In general, we provide significant evidence of a role for judicial ideology in judicial review of agency decisions for arbitrariness. A key goal of the arbitrary and capricious standard is to ensure that judges invalidate agency actions when those actions reflect serious analytic errors or palpable political pressures, and to prevent these errors and pressures from being translated into grounds for law.[45] Most ambitiously, arbitrariness review can be seen as a response to the uneasy constitutional position of agencies wielding broad discretionary power; perhaps such review can reintroduce surrogate safeguards for the decline of constitutional checks on agency authority. But if Democratic appointees are especially inclined to find conservative decisions to be arbitrary, and if Republican appointees are especially likely to find liberal decisions to be arbitrary, something is seriously amiss.

Notably, the role of political judgments appears to be strikingly similar when courts are reviewing agency interpretations of law under *Chevron U.S.A. Inc v NRDC*[46] and when judges are addressing questions of fact and policy under arbitrariness review. The numbers are very close in the two contexts. This finding suggests that at least in the domain of EPA and NLRB decisions, ideology influences judges' decisionmaking to the same extent regardless of the judicial task or the standard of review. Moreover, the degree of ideological influence seems roughly the same for both tasks and under both standards.

Our findings offer a clear prediction for the future: when a judiciary consisting mostly of Democratic appointees confronts a conservative executive branch, the rate of invalidations will be unusually high, and so too when a judiciary consisting mostly of Republican appointees confronts a liberal executive branch. The conflict between a Democratic administration and a Republican-dominated judicial branch should be expected to produce a large number of invalidations in the most important domains of regulatory policy. Notably, such invalidations will typically involve complex questions of law and fact not readily suited to

---

43 See Miles and Sunstein, 73 U Chi L Rev at 849 (cited in note 39) (finding that circuit courts have a 64 percent validation rate when evaluating agency decisions under *Chevron*).

44 See Breyer, 38 Admin L Rev at 397 (cited in note 10).

45 See Merrick B. Garland, *Deregulation and Judicial Review*, 98 Harv L Rev 505, 553–56 (1985) (contending that the role of hard look review in deregulation cases is to "ferret out" true irrationality or illegitimate motives).

46 467 US 837 (1984).

oversight by the Supreme Court. We will offer some suggestions about how existing doctrines might change to counteract the evident risks.

I. ADMINISTRATIVE LAW PRELIMINARIES

To understand our study, some background is in order. Agency decisions might be challenged on many possible grounds. Most obviously, their decisions might violate a governing statute. With respect to such challenges, much of the current doctrine is organized under the framework established by the Court's *Chevron* decision.[47] That decision provides a famous two-step test for evaluating agency interpretations of law. The first question is whether the agency has violated an unambiguous provision of law. If not, the court proceeds to the second question, which is whether the agency's interpretation of an ambiguous provision is reasonable.[48] But many agency interpretations are not entitled to judicial deference under *Chevron*, and such interpretations will receive less deference, or even no deference, from reviewing courts.[49] We are not concerned with agency interpretations of law here, except by way of comparison.

Agency decisions might also be challenged as inconsistent with the procedural requirements of the APA[50] or any other applicable statute. For example, the agency might have issued a rule without using notice and comment procedures,[51] or it might have violated a statutory prohibition on ex parte communications.[52] Judicial review of agency compliance with the APA's procedural requirements raises many important questions, and an empirical study of the relevant judicial decisions would undoubtedly be instructive. Perhaps ideological voting can be found in this domain as well. Might Democratic appointees be comparatively more willing to find violations of the procedural requirements of the APA when the agency has issued a conservative rule? Might Republican appointees be more willing to in-

47 See id at 842–44.

48 Id at 843–44.

49 See *United States v Mead Corp*, 533 US 218, 226–27 (2001) (holding that *Chevron* deference applies "when it appears that Congress delegated authority to the agency generally to make rules carrying the force of law, and that the agency interpretation claiming deference was promulgated in the exercise of that authority"); Cass R. Sunstein, Chevron *Step Zero*, 92 Va L Rev 187, 193 (2006) (describing a line of cases suggesting that *Chevron* may not apply if Congress has not delegated the power to act with the force of law and a different line of cases suggesting that *Chevron* does not apply if a "fundamental issue" is at stake).

50 5 USC § 551 et seq (2000).

51 See, for example, *Community Nutrition Institute v Young*, 818 F2d 943, 948–49 (DC Cir 1987) (holding that failure to comply with the notice and comment procedure invalidates agency action).

52 See, for example, *Professional Air Traffic Controllers Organization v Federal Labor Relations Authority*, 685 F2d 547, 574 (DC Cir 1982) (finding that there was not sufficient evidence of ex parte communications to "irrevocably taint[]" the agency decision and require remand).

validate agency decisions on procedural grounds when those decisions turn out to be liberal? We suspect so, and it would be valuable to know; but we do not explore such issues here.

Our focus is on the question of whether agency decisions are unlawful because they are arbitrary or lack substantial evidence. That question might be understood as a kind of Step Three, to be asked directly after the two-step inquiry mandated by *Chevron*. To understand that question, it is necessary to explore the hard look doctrine and *State Farm* in somewhat more detail.

The case itself involved an important controversy over the legal validity of a change in regulatory policy initiated by the Reagan Administration. Under President Carter, the National Highway Traffic Safety Administration (NHTSA) had shown considerable interest in "passive restraints"—in the form of automatic seatbelts or airbags—that would protect drivers even if they failed to take action to buckle up. The ultimate regulation, issued in the closing months of the Carter Administration, required automobile manufacturers to equip new cars with one of three possible passive restraints: detachable seat belts, nondetachable seat belts, or air bags. NHTSA concluded that the new rule would produce at least a 13 percent increase in seat belt usage and that, as a result, its benefits would justify its costs.[53]

Within six months, President Reagan's NHTSA repealed the regulation.[54] In brief, the agency concluded that, contrary to the analysis conducted under the Carter Administration, the regulation would not produce a significant increase in seat belt usage, and hence the benefits were too speculative to justify the imposition of the passive restraints rule on manufacturers.[55] This conclusion was challenged as arbitrary; the challengers invoked the hard look doctrine.[56] The government responded quite ambitiously, by attacking that doctrine as illegitimate; in its view, agency action must be upheld unless it was found to be wholly irrational through a highly deferential analysis similar to that undertaken under the Due Process Clause.[57] Notwithstanding its ambition, this argument did not seem implausible in light of the Court's then-recent decision in *Vermont Yankee Nuclear Power Corp v NRDC*,[58] which emphasized that judges had no business burdening agencies with duties that could not be found in the APA or

53 See *State Farm*, 463 US 29, 35–39, 51–55 (1983).
54 Id at 38–40.
55 Id at 54–55.
56 Id at 40.
57 Id at 43 n 9.
58 435 US 519 (1978).

some other source of law.[59] Moreover, the government's objection to the hard look doctrine could draw strength from the view, pressed by many skeptics in the period, that liberal judges had used the doctrine to push regulatory policies in the directions that they themselves favored on the merits.[60]

In striking down the repeal of the regulation, the Court endorsed both procedural and substantive aspects of the hard look doctrine. Speaking in general terms, the Court unanimously said that under arbitrariness review, agencies must provide detailed explanations of their action, offer careful attention to counterarguments, and show serious engagement with alternatives.[61] On the merits, the Court concluded, again by a unanimous vote, that the repeal of the regulation was arbitrary because NHTSA had not investigated whether an "airbags only" alternative would have produced sufficient benefits to justify the rule.[62] By a vote of 5-4, the Court also held that the agency's analysis of detachable and nondetachable seat belts was arbitrary because it depended on unsupported judgments about likely facts.[63]

The 5-4 division within the Court is especially noteworthy for our purposes, for it occurred along evidently political lines. The dissenting opinion was written by then-Justice Rehnquist, who emphasized that it was entirely appropriate for President Reagan to reject the policies of his predecessor. In his words,

> [t]he agency's changed view of the standard seems to be related to the election of a new President of a different political party. . . . A change in administration brought about by the people casting their votes is a perfectly reasonable basis for an executive agency's reappraisal of the costs and benefits of its programs and regulations.[64]

Thus, Justice Rehnquist offered a firm plea for judicial deference in the face of the new commitments of a new administration—a plea that bears directly on the data we shall offer here.

---

59 See id at 550–51 ("Time and resources are simply too limited to hold that an impact statement fails because the agency failed to ferret out every possible alternative.").

60 See, for example, Melnick, *Regulation and the Courts* at 12 (cited in note 31).

61 See *State Farm*, 463 US at 47–51 (affirming that an agency must explain the reasoning behind its decisions, not as a matter of procedure but in order to demonstrate its thorough consideration of the issue at hand).

62 See id at 48–49 (indicating that the agency had described air bags as effective and yet did not consider the viable alternative of mandatory air bag installation).

63 See id at 51–55 (finding that agency reasoning that equates the effects of detachable and nondetachable seat belts is inadequate to refute specific determinations made by the previous administration).

64 Id at 59 (Rehnquist concurring in part and dissenting in part).

State Farm was widely taken to have ratified the hard look doctrine.[65] The Court's description of the appropriate standard of review, and its conclusions on the merits, suggested that courts should require detailed justifications for agency action and also examine the reasonableness of the agency's conclusions. There was obvious tension, however, between *State Farm* and *Chevron*, which was decided just one year later.[66] Under *State Farm*, courts would take a hard look at agency judgments of policy and fact; under *Chevron*, courts would give considerable deference to agency interpretations of ambiguous statutes. Hence it was natural to object, as did then-Judge Breyer, that a sensible system of judicial review would not entitle courts to give careful scrutiny to judgments of policy and fact while also requiring them to defer on questions of law.[67] If we attend to the distinctive competence of agencies and courts, the opposite conclusion might seem hard to resist: questions of law are for judicial resolution, whereas questions of policy and fact should be resolved by agencies.

But this simple comparison between *State Farm* and *Chevron* misses some complexities. First, it may not be correct to suggest that the former decision implies "less" deference than the latter. Under *Chevron*, agencies must obey unambiguous statutes,[68] and even when there is ambiguity, agency interpretations must be reasonable.[69] Under *State Farm*, agency decisions will also be upheld so long as they are reasonable.[70] In the abstract, it would be possible to read the two rulings in a way that would not create the anomaly to which Justice Breyer objects. In any case, it is much too simple to suggest that courts

65 See, for example, Sunstein, 1983 S Ct Rev at 178 (cited in note 20) (claiming that the Court endorsed aspects of the hard look doctrine developed by the DC Circuit); Breyer, 38 Admin L Rev at 387–88 (cited in note 10) (noting that the *State Farm* Court stated that it applied the arbitrary and capricious standard but actually applied "strict review" earlier described as the hard look doctrine); Peter L. Strauss, *Considering Political Alternatives to "Hard Look" Review*, 1989 Duke L J 538, 539 (remarking that the *State Farm* decision "essentially endorsed" the hard look standard); Kathryn A. Watts, *Adapting to Administrative Law's* Erie *Doctrine*, 101 Nw U L Rev 997, 1043 n 273 (2007) (stating that the hard look doctrine is derived from *State Farm*).

66 See Breyer, 38 Admin L Rev at 373, 387–88 (cited in note 10).

67 See id.

68 See, for example, *Public Citizen v Young*, 831 F2d 1108, 1122 (DC Cir 1987) (finding that the FDA's construction of a statute violated clear congressional intent and thus was not entitled to respect).

69 See, for example, *Ohio v Department of the Interior*, 880 F2d 432, 464 (DC Cir 1989) (holding that the Department of the Interior's refusal to account for option and existence values was not a reasonable interpretation of the statute).

70 See, for example, *Syracuse Peace Council v FCC*, 867 F2d 654, 669 (DC Cir 1989) (holding that the agency's decision to rescind the "fairness doctrine"—and thus its requirements regarding broadcast media coverage of vital public issues—was reasonable).

should decide questions of law on their own.[71] Where statutes are ambiguous, the resolution of the ambiguity frequently requires judgments of policy and principle. *Chevron* rests on the belief that such judgments should be made by officials with a degree of accountability and specialized competence, not by judges.[72]

But if this point is correct, *State Farm* itself must be taken with some caution. Review of agency decisions for arbitrariness often involves highly technical issues of policy and fact, and rulings by courts of appeals are usually too particularistic to be well suited to Supreme Court review. If *State Farm* is operating in a way that reflects judicial policy preferences, Justice Breyer's objection has considerable force.

The empirical questions emerge as the important ones. What, exactly, have appellate courts[73] been doing? What might be said about the real world of arbitrariness review? It is to these questions that we now turn.

II. ARBITRARINESS REVIEW OF EPA AND NLRB DECISIONS IN THE COURTS OF APPEALS

A. Data and Method

We devote our attention to two agencies whose decisions have a high degree of practical importance and political salience: the EPA and the NLRB.[74] We extracted from the standard legal databases a list of appellate court cases that applied arbitrary and capricious or sub-

---

71 See, for example, E. Donald Elliott, Chevron *Matters: How the* Chevron *Doctrine Redefined the Roles of Congress, Courts and Agencies in Environmental Law*, 16 Vill Envir L J 1, 16–18 (2005) (arguing that judicial deference to agency interpretations increases democratic legitimacy because decisions rest on "science-based determinations that Congress authorized" by agencies subject to executive oversight, not on one judge's interpretation of the congressional history, which often serves to mask the judge's policy preferences); Antonin Scalia, *Judicial Deference to Administrative Interpretations of Law*, 1989 Duke L J 511, 513 (citing the Attorney General's Administrative Procedure Report as stating that "[e]ven on questions of law [independent judicial] judgment seems not to be compelled" and that agency interpretations should play a powerful role).

72 See Cass R. Sunstein, *Beyond* Marbury*: The Executive's Power to Say What the Law Is*, 115 Yale L J 2580, 2587 (2006) (contending that *Chevron* rests on the assertion that agencies are better equipped for statutory interpretation because of agency expertise, oversight by the executive, and the ability to more quickly and adequately adapt statutes to new situations).

73 We put decisions by the Supreme Court to one side, on the ground that the Court has decided very few cases involving "arbitrariness" review, and hence the small number of observations would make it difficult to draw inferences.

74 A valuable discussion of judicial review of NLRB decisions, involving a different data set and somewhat different questions but overlapping conclusions, is James J. Brudney, Sara Schiavoni, and Deborah J. Merritt, *Judicial Hostility toward Labor Unions? Applying the Social Background Model to a Celebrated Concern*, 60 Ohio St L J 1675 (1999) (studying the impact of educational experience, race, sex, political affiliation, and work experience on circuit court judges' decisions in cases involving the National Labor Relations Act, and controlling for deference to the NLRB).

stantial evidence review to decisions of the EPA and the NLRB[75] between 1996 and 2006. There were 653 cases in all, and a strong majority of these cases (554) reviewed NLRB decisions. It would be natural to think that in view of these numbers, our focus is largely on review of NLRB decisions, but most of the patterns do not significantly differ as between review of EPA decisions and review of NLRB decisions. Indeed, we have studied cases citing *State Farm* itself, in which EPA cases form a majority;[76] in that data set, the EPA cases are more numerous than the NLRB cases, and many of the patterns we find here—above all, a significant ideological split between Democratic and Republican appointees—can be found there as well. Where the differences are significant, we report them. (Our strong suspicion is that the same general patterns would be found for other agencies, but that point must remain speculative for now.) We coded the votes of the individual judges in these cases and assembled a file of judge-by-case observations. Of the 1,959 total votes by judges in these cases, 807 were from appointees of Democratic presidents and 1,152 were from appointees of Republican presidents.

For all of the key questions, illuminating patterns emerge, allowing us to assess party and panel effects in arbitrariness review. We are also able to disaggregate the data in such a way as to cast light on questions that have been explored in the literature on both arbitrariness review and judicial behavior under *Chevron*.[77]

To test the role of judicial policy judgments, we use several interacting measures. For judges, we focus on the party of the appointing president, because that factor has importance in its own right and because it serves as a rough proxy for the ideological preferences of the judges. To say the least, it would be valuable to know if Democratic appointees are especially likely to find arbitrariness on the part of Republican administrations, or if the validation rates of Republican appointees increase when the president is a Republican. Although political scientists have legitimately criticized the use of party as a proxy for political ideology[78] and often use "common space scores" as

---

75 See note 39.

76 See Thomas J. Miles and Cass R. Sunstein, *The Hard Look in Practice* 5–6, 10 (unpublished manuscript, 2007).

77 We conducted regression analysis to verify the robustness of our findings to other possible influences on a judge's review of an arbitrariness challenge. In addition to the variables described in Parts II and III, the regressions included fixed effect controls for circuits and years. The regression analysis did not alter the conclusions we reached from examining summary statistics, and therefore, we omit the regressions from our discussion here.

78 See, for example, Lee Epstein and Gary King, *The Rules of Inference*, 69 U Chi L Rev 1, 88–89 (2002). See also Joshua B. Fischman, *Decision-making under a Norm of Consensus: A Structural Analysis of Three-judge Panels* 2–3 (First Annual Conference on Empirical Legal Stud-

an alternate measure,[79] the political party of the appointing president remains a valuable tool of inquiry,[80] especially for those interested in the power of the executive to move the federal judiciary in its preferred directions.

We measure the political content of agency decisions in two distinct ways. First, we classify agency decisions as "conservative" or "liberal" on the basis of the identity of the party making the challenge. When a labor union or public interest group challenges an agency decision, we deem it to be relevantly "conservative." When an industry group or corporation challenges the agency's decision, we code it as relevantly "liberal." The reason for this approach is that the reviewing court assesses the position of an agency not in the abstract but in relation to the claims of the particular challenger. When a public interest group, such as the Sierra Club or the Natural Resources Defense Council, brings a challenge, the agency appears conservative relative to the challenger. When a corporation challenges an agency decision that regulates water pollution or finds an unfair labor practice, the agency appears liberal relative to the challenger. This coding scheme does introduce some imprecision, which we attempted to correct by investigating individual cases; but as compared to the alternative of ad hoc evaluation of each agency policy, its objectivity and its easy administrability are its virtues. It is important not to be confused by the measure: a Republican administration might issue many decisions that are relevantly liberal, in the sense that companies find it worthwhile to challenge them, and observers might nonetheless conclude that those decisions are conservative by some objective measure.

Our second proxy for the political direction of the agency decision is whether the case was decided during a Democratic or Republican administration. As a general rule, the positions of agencies under Democratic presidencies are certainly more liberal (or less conservative) than those of agencies under Republican presidencies. To the

---

ies, 2006) (positing a structural model that allows for unobserved heterogeneity within parties of the appointing president).

79 See, for example, Gregory C. Sisk and Michael Heise, *Judges and Ideology: Public and Academic Debates about Statistical Measures*, 99 Nw U L Rev 743, 784–85 (2005) (describing how a "common space score" accounts for variation within a political party and for instances when the president defers to senators' recommendations for judicial appointment); Keith T. Poole and Howard Rosenthal, *Congress: A Political-economic History of Roll Call Voting* 11–26 (Oxford 1997) (describing a multidimensional empirical model of congressional roll call voting occupying Euclidean space); Nolan M. McCarty and 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, 288 (1995) (employing a one-dimensional spatial model of policy preferences).

80 See Daniel R. Pinello, *Linking Party to Judicial Ideology in American Courts: A Meta-analysis*, 20 Just Sys J 219, 243 (1999) (analyzing eighty-four empirical studies and finding that there is a link between political party affiliation and judicial voting behavior).

extent that this generalization is crude, it remains independently important to understand how judicial behavior changes across administrations. A potential difficulty with this measure is that litigation may take years to resolve, and courts of appeals might well be asked to evaluate a regulation, initially issued under the Clinton Administration, during the Bush Administration. When litigation extends across administrations of opposing parties, misattributions may occur. But it is not entirely clear that the relevant question is the political affiliation of the administration that initially issued a regulation or a final order; perhaps what matters is the affiliation of the administration that is litigating the case. Note that a new administration has the opportunity to reverse agency positions and to settle ongoing cases before the court issues its decision.[81] In any event, our findings are not significantly affected if we adjust the data to consider the administration that originally issued the regulation or order.

B. Judicial Votes and Partisan Affiliations

1. Validation rates.

Table 1 reports the rates at which appellate judges vote to validate the decisions of the EPA and the NLRB under the arbitrary and capricious and substantial evidence standards.[82] Column (1) shows the total validation rates for Democratic and Republican appointees. It reveals that overall, Democratic appointees vote with significantly higher frequency to validate decisions of the EPA and the NLRB. The total validation rate under arbitrariness review is quite close to the validation rate under *Chevron* of 64 percent.[83] But a difference between arbitrariness review and *Chevron* is immediately apparent. In the *Chevron* cases, the overall validation rates of Democratic and Republican appointees are the same, while under arbitrariness review the validation rate of Democratic appointees is 10 percentage points higher than that of Republican appointees.

---

81 For ease of exposition, we refer to the administrations as Republican or Democratic, but over the time period studied, 1996–2006, the Clinton Administration was the only Democratic administration, and the administration of George W. Bush was the only Republican administration.

82 It is generally believed that EPA decisions are typically reviewed under the arbitrary and capricious standard, and NLRB decisions under the substantial evidence standard, but the belief is too crude. Some NLRB decisions are set aside as arbitrary, and some EPA decisions are evaluated, under relevant statutes, for lack of substantial evidence. See, for example, *Islander East Pipeline Co v Connecticut Department of Environmental Protection*, 482 F3d 79, 100 (2d Cir 2006) (holding that the EPA's decision that a pipeline would dramatically alter the surrounding habitat was not supported by substantial evidence); *New England Health Care Employees Union v NLRB*, 448 F3d 189, 193 (2d Cir 2006) (holding that the NLRB's determination that an employer did not act with unlawful intent was arbitrary and capricious).

83 See Miles and Sunstein, 73 U Chi L Rev at 849 (cited in note 39).

TABLE 1
VALIDATION RATES OF CIRCUIT COURT JUDGES IN ARBITRARINESS REVIEW CASES BY PARTY OF APPOINTING PRESIDENT AND BY IDEOLOGICAL CONTENT OF AGENCY DECISION

	Ideological content of agency decision			
Party of appointing president	Total (1)	Liberal (2)	Not liberal (3)	Difference of (2)–(3)
(A) Democrat	0.699 (0.016) [807]	0.721 (0.017) [703]	0.548 (0.049) [104]	0.173*** (0.048)
(B) Republican	0.596 (0.014) [1,152]	0.582 (0.015) [1,037]	0.722 (0.042) [115]	-0.139** (0.048)
Difference of (A)–(B):	0.103*** (0.022)	0.139** (0.023)	-0.174** (0.064)	

Means, standard errors in parentheses, and number of observations in brackets.
Note: * denotes difference significant at 10 percent level, ** denotes difference significant at 5 percent level, and *** denotes difference significant at 1 percent level. Differences may not match exactly due to rounding.

What accounts for this difference? The answer lies in the higher proportion of liberal agency decisions among the arbitrariness cases, which makes ideological differences immediately detectable. Decisions by the NLRB comprise a far larger share of our data set here, and most of these decisions are liberal by our standards. (Recall that those standards are relative, not absolute; an employer might challenge an NLRB decision that has a Republican majority, even though the decision is far more conservative than what would emerge from an NLRB decision with a Democratic majority.) NLRB decisions account for 85 percent of the arbitrariness cases but only 28 percent of the *Chevron* cases.[84] In addition, over 94 percent of the NLRB decisions reviewed for arbitrariness or lack of substantial evidence were liberal,[85] while 67 percent of those reviewed under *Chevron* were coded as liberal. In contrast, EPA decisions, under both *Chevron* and arbitrariness review, were roughly evenly split between liberal and conservative decisions. Because most agency decisions subjected to arbi-

84 See id at 825.

85 This figure is consistent with prior studies of NLRB decisionmaking. See, for example, William N. Cooke, et al, *The Determinants of NLRB Decision-making Revisited*, 48 Indus Labor Rel Rev 237, 239 (1995) (showing that nearly 87 percent of unfair labor practice disputes that remained unresolved prior to a circuit court or Supreme Court decision were brought against employers).

trariness review were liberal, ideological differences are immediately apparent in the arbitrariness data.

Although we do not have a definitive explanation for the differences in the number and nature of the two agencies' decisions, we believe that the underlying reason is straightforward.[86] The EPA makes essentially all of its policies via rulemaking, and the number of rules in any particular year is relatively small. Because they are conducted against the background of clear regulations, EPA adjudications frequently involve the application of settled law to not-much-disputed fact, and the room for challenge in court is not large. By contrast, the NLRB makes essentially all of its policies via adjudication, and the number of adjudications in any particular year is large. Moreover, there is considerable room for challenging the NLRB's judgments about policy and fact. The making of national labor policy through case-by-case decisionmaking has attracted considerable skeptical attention within the Supreme Court itself.[87] The large number of NLRB cases in our sample reflects the fact that the NLRB makes many more decisions that are subject to a plausible challenge on arbitrariness or substantial evidence grounds. By contrast, a high percentage of EPA decisions can be challenged on *Chevron* grounds, and a high percentage of NLRB decisions cannot be; hence, the proportions in *Chevron* cases are not so lopsided.

This sizable gap in the ideological direction of the agency decisions generates a difference in the validation rates across the two agencies. Figure 1 shows the rates at which judges of both parties voted to validate the decisions of the two agencies. The EPA enjoyed a higher rate of validation when its decisions faced challenges for arbitrariness. Overall, judges voted to validate EPA decisions 72 percent

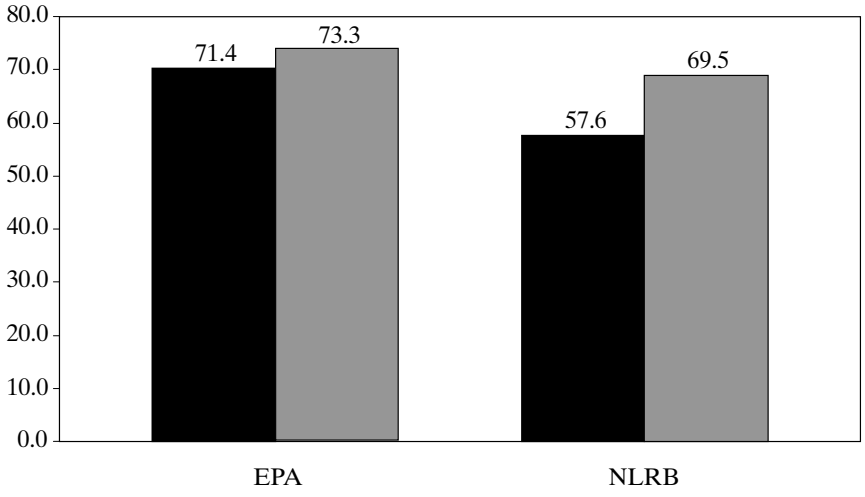
---

86 Researchers have long observed wide differences across agencies in the rate at which the Supreme Court validates their decisions. Bradley C. Canon and Micheal Giles, *Recurring Litigants: Federal Agencies before the Supreme Court*, 25 W Polit Q 183, 184 (1972) (reporting that agency success rates before the Court range from 56 percent to 91 percent); Roger Handberg, *The Supreme Court and Administrative Agencies, 1965–1978*, 6 J Contemp L 161, 168 (1979) (extending the Canon and Giles data set's scope by five years and reporting agency success rates before the Supreme Court ranging from 55 percent to 91 percent); Donald W. Crowley, *Judicial Review of Administrative Agencies: Does the Type of Agency Matter?*, 40 W Polit Q 265, 271 (1987) (finding that agencies classified as economic have a 79 percent success rate before the Court while those classified as social have a 68 percent success rate); Reginald S. Sheehan, *Administrative Agencies and the Court: A Reexamination of the Impact of Agency Type on Decisional Outcomes*, 43 W Polit Q 875, 880 (1990) (reporting that social and economic agencies have similar success rates overall but that substantial variation exists when social agencies' decisions are classified as liberal or conservative).

87 See *Allentown Mack & Service, Inc v NLRB*, 522 US 359, 372–73 (1998) (objecting that the NLRB may use a standard other than the "announced standard" in its adjudications).

of the time compared to only 62 percent for decisions of the NLRB.[88] EPA decisions also produced a more modest partisan gap in the judges' voting. The rate at which Democratic and Republican appointees voted to reject arbitrariness challenges to EPA decisions differed by fewer than 2 percentage points. The gap for NLRB decisions was considerably larger. Democratic appointees voted to validate NLRB decisions 70 percent of the time, while Republican appointees did so only 58 percent of the time.

FIGURE 1
VALIDATION RATES OF CIRCUIT COURT JUDGES IN ARBITRARINESS REVIEW CASES BY AGENCY AND BY PARTY OF APPOINTING PRESIDENT



Note: The darkly shaded bars indicate the validation rates of Republican appointees, and the lightly shaded bars indicate the validation rates of Democratic appointees.

The differences across agencies have many similarities as well as some contrasts with our earlier findings with respect to *Chevron* review.[89] The primary contrast is that under *Chevron*, the NLRB enjoyed

[88] Despite the difference across these two agencies, these validation rates are similar to those prior researchers have found in appellate court review of administrative agency decisions. See Martha Anne Humphries and Donald R. Songer, *Law and Politics in Judicial Oversight of Federal Administrative Agencies*, 61 J Polit 207, 215 (1999) (finding that the approval rate for all agency decisions was 58 percent); David H. Willison, *Judicial Review of Administrative Decisions: Agency Cases before the Court of Appeals for the District of Columbia, 1981–1984*, 14 Am Polit Rsrch 317, 321 (1986) (finding that the approval rate for EPA decisions was 59 percent, the approval rate for NLRB decisions was 64 percent, and the approval rate for all agency decisions was 66 percent).

[89] For earlier findings, see Miles and Sunstein, 73 U Chi L Rev at 852–54 (cited in note 39).

a slightly higher validation rate than the EPA, while under arbitrariness review, this ordering is reversed. The primary common feature is that under both arbitrariness review and *Chevron*, the partisan gap in validation rates is largest for NLRB decisions. Evidently the labor-management relations that come to the federal courts of appeals are more ideologically contentious than are the environmental issues, which might well appear more technical.

When we decompose the data by examining the ideological content of the agency decisions, we find even more substantial differences in the behavior of the two groups of appointees.

2. Political voting and failed aspirations.

Columns (2) and (3) of Table 1 stratify the voting rates by the partisan policy direction of the agency decision, and we now see an especially sharp contrast in the voting patterns of Republican and Democratic appointees. We are able to demonstrate for the first time[90] that judicial policy judgments play an unquestionable role under arbitrariness review.

a) Liberal agencies, conservative agencies. When the agency decision is conservative, Democratic appointees conclude that the decision was not arbitrary or capricious at a rate that is 17 percentage points lower than when the agency decision is liberal. The pattern is in the opposite direction for Republican appointees. When the agency decision is liberal, Republican appointees conclude that the decision is not arbitrary or capricious at a rate that is 14 percentage points lower than when the agency decision is conservative. These patterns imply that the validation rates of Democratic appointees are nearly 14 percentage points above those of their Republican counterparts for liberal agency decisions and about 17 percentage points below those of Republican appointees for conservative agency decisions. To say the least, this is a dramatic difference in the operation of hard look review.

These findings contain striking similarities to our previous analysis of judicial review under *Chevron*. The frequency of agency validation is nearly identical under arbitrary and capricious review as it is under *Chevron*; both are about 64 percent.[91] This finding bears on the concern voiced by Justice Breyer two decades ago, to the effect that under existing doctrine, agencies might be significantly more likely to

90 A more limited data set, focusing on EPA decisions in the DC Circuit, reaches a similar conclusion. See Revesz, 83 Va L Rev at 1719 (cited in note 34) (asserting that "ideology significantly influences decisionmaking on the D.C. Circuit").

91 Miles and Sunstein, 73 U Chi L Rev at 849 (cited in note 39).

lose on issues of fact and policies than on issues of law.[92] Our data do not confirm this prediction. Rather, the data appear consistent with now-standard analyses of litigant decisionmaking. Litigants should be expected to adjust their behavior to the prevailing standard of review,[93] and the roughly similar validation rates under arbitrariness review and *Chevron* suggest that litigants readily make these adjustments. Because litigants are likely to adjust their decisions in accordance with the intensity of review, our figures cannot be taken to answer the question of whether *Chevron* review is more rigorous than arbitrariness review, or vice versa. A constant rate of 64 percent is possible even if one standard is far more searching than another—at least if we assume, as seems likely, that the selection of cases will be affected by litigant perceptions of when they are least likely to lose.

Both *Chevron* and *State Farm* seek to cabin the influence of judicial ideology in review of agency decisionmaking. An evident aspiration of the *Chevron* approach is to limit the role of judicial judgments in the domain of policy.[94] Despite its command of deference to reasonable agency interpretations of law, the persistence of judicial politics under *Chevron* is plain.[95] *State Farm* does call for judicial scrutiny of agency judgments about fact and policy, but the Court made clear that so long as the agency offered "a reasoned analysis," it would be permitted to do as it saw fit.[96] Indeed, *State Farm* must be taken in the context of both *Citizens to Preserve Overton Park, Inc v Volpe*,[97] where the Court emphasized that "the ultimate standard of review is a narrow one" affording the agency "a presumption of regularity,"[98] and *Vermont Yankee*, where the Court stressed that the ultimate decision is for agencies, not for courts.[99] The Court has yet to offer an unambiguous warning about the politicization of judicial review under scrutiny of possible "arbitrariness," but the key decisions are plainly meant to reduce the relevant risks.

---

92 See Breyer, 38 Admin L Rev at 397 (cited in note 10).

93 See George L. Priest and Benjamin Klein, *The Selection of Disputes for Litigation*, 13 J Legal Stud 1, 5 (1984) (arguing that "the individual maximizing decisions of the parties will create a strong bias toward a rate of success . . . of 50 percent regardless of the substantive standard of law"). See also Part IV.A.

94 See Elliott, 16 Vill Envir L J at 18 (cited in note 71).

95 See Miles and Sunstein, 73 U Chi L Rev at 860 (cited in note 39).

96 463 US at 42.

97 401 US 402 (1971).

98 See id at 415–16 (explicating a standard that is less than substantial evidence but is still a "thorough, probing, in-depth review").

99 See 435 US at 524 ("[T]his Court has for more than four decades emphasized that the formulation of procedures was basically to be left within the discretion of the agencies to which Congress had confided the responsibility for substantive judgments.").

Notwithstanding the Court's aspirations, the figures in Table 1 show a large role for judicial policy preferences in the operation of arbitrariness review. The magnitude of the fluctuation in validation rates between liberal and conservative agency decisions is roughly the same in arbitrariness review cases as it is in *Chevron* cases.[100] Under both standards, the validation rates of Democratic and Republican appointees seesaw in response to the ideological content of the agency decision. When the agency decision is liberal, Democratic appointees validate more often than Republican appointees by about 14 percentage points, and when the agency decision is conservative, Republican appointees validate more often by about 17 percentage points. The results demonstrate that arbitrariness review under the *State Farm* framework has failed to eliminate the influence of judicial ideology in review of agency decisions of policy and fact.[101]

(b) Republican administrations, Democratic administrations. Table 2 presents validation rates for the two groups of judges when the party of the current president is used as the measure of the political valence of the agency decision. We anticipated that Republican appointees would be more likely to uphold decisions under Republican administrations and that Democratic appointees would be more likely to uphold decisions under Democratic administrations—and indeed this is the pattern we observe in the *Chevron* context.[102]

We were initially surprised to find that for Democratic appointees, the validation rates do not correlate with the party of the current president in the predicted ways. The validation rates of Democratic appointees do not vary with the party of the current president. By contrast, those of Republican appointees do move in the predicted direction. During a Democratic administration, Democratic appointees vote to validate EPA and NLRB decisions 70 percent of the time, a rate virtually identical to their validation rate during Republican administrations. But the validation rates of Republican appointees are about 8 percentage points higher during Republican administrations.

---

100 See Miles and Sunstein, 73 U Chi L Rev at 849 (cited in note 39) (reporting that Democratic appointees validate liberal agency decisions more often than Republican appointees by about 14 percentage points and that Republican appointees validate conservative agency decisions more often by about 19 percentage points).

101 When we decompose these data by individual agency, the patterns are similar. The seesaw pattern is very sharp in the NLRB decisions. It is more muted in the EPA decisions. There, the validation rates of Republican appointees respond little to the ideological direction of the agency decision whereas the validation rates of Democratic appointees fluctuate by about 20 percentage points.

102 See Miles and Sunstein, 73 U Chi L Rev at 850 (cited in note 39).

TABLE 2
VALIDATION RATES OF CIRCUIT COURT JUDGES IN ARBITRARINESS REVIEW CASES BY PARTY OF APPOINTING PRESIDENT AND BY PARTY OF CURRENT PRESIDENT

	Party of current president			
Party of appointing president	Total (1)	Democrat (2)	Republican (3)	Difference of (2)–(3)
(A) Democrat	0.699 (0.016) [807]	0.698 (0.022) [427]	0.700 (0.025) [380]	-0.002 (0.032)
(B) Republican	0.596 (0.014) [1,152]	0.561 (0.020) [617]	0.637 (0.021) [535]	-0.077** (0.029)
Difference of (A)–(B):	0.103*** (0.022)	0.137*** (0.030)	0.063** (0.032)	

Means, standard errors in parentheses, and number of observations in brackets.
Note: * denotes difference significant at 10 percent level, ** denotes difference significant at 5 percent level, and *** denotes difference significant at 1 percent level. Differences may not match exactly due to rounding.

These patterns show that when reviewing EPA and NLRB decisions for arbitrariness, the validation rates of Democratic appointees are higher than those of Republican appointees, irrespective of which party currently holds the Presidency. During Democratic Presidencies, the validation rates of Democratic appointees in these cases were 14 percentage points higher than that of Republican appointees, and during Republican Presidencies, this difference was only 6 percentage points. But the gap between these figures (14 versus 6) is not statistically meaningful. What is clear is that Democratic appointees validate EPA and NLRB decisions at higher rates during both Democratic and Republican administrations during this period.

At first blush, Tables 1 and 2 present a confusing and inconsistent picture of whether judges are responsive to the ideological content of the agency decision. In Table 1, the validation rates of appointees from both parties appear highly responsive to political considerations, while in Table 2, the validation rates of Democratic appointees appear consistently higher than those of Republican appointees. How can these patterns be explained?

The answer lies in the fact that Republican administrations often produce a rule, decision, or order that is relevantly liberal, in the sense that it is challenged by a company that is regulated by an EPA rule or displeased by a finding of an unfair labor practice by the NLRB. So too, a Democratic administration may and often does produce a deci-

sion or rule that is relevantly conservative, in the sense that it is challenged by a public interest group or a labor union. It is for this reason that in these data, the purely political coding of the agency decision is a far more accurate measure than the party of the administration at the time the court issues its ruling. As previously described, our data set here consists disproportionately of liberal decisions by the NLRB.[103]

For that reason, it should not be puzzling that in Table 2 the validation rates of Democratic appointees appear higher than those of Republican appointees in both Republican and Democratic administrations. What matters most is whether the agency decision was liberal, not whether it was issued under a Republican president. It follows that if the goal is to assess the role of judicial ideology in arbitrariness review, there is reason for much greater confidence in the estimates based on our direct coding of the agency decisions.

With these qualifications in mind, the central findings are clear. In cases applying arbitrariness review, the validation rates of Democratic appointees exceed those of Republican appointees by about 10 percentage points. When individual agency decisions are classified according to their ideological content, the role of politics is unmistakable: Democratic appointees are far more likely to uphold liberal decisions than conservative ones, and Republican appointees show the opposite pattern. Arbitrariness review is being applied in a way that shows a large influence from judicial policy preferences.

C. Panel Effects

1. The standard patterns.

A great deal of evidence shows that the composition of appellate panels significantly influences the voting behavior of individual judges. In many domains, the standard pattern includes both *ideological dampening* and *ideological amplification*.[104] Dampening occurs when a Democratic appointee shows unusually conservative voting patterns if sitting with two Republican appointees, and when a Republican appointee shows unusually liberal voting patterns if sitting with two Democratic appointees. Amplification occurs when the most liberal

---

[103] A comparison of the number of conservative agency decisions in Table 1 with the number of decisions arising during Republican administrations in Table 2 makes this point clear. The total number of judges' votes in conservative agency decisions, 219 (104 + 115), is far less than the total number of judges' votes during Republican administrations, 915 (380 + 535). These totals imply that under the Republican administration during our observation period, most EPA and NLRB decisions reviewed for arbitrariness by Republican appointees were "liberal" in our sense, and this finding explains the apparent anomalies in the voting patterns of those appointees.

[104] See Cass R. Sunstein, et al, *Are Judges Political?: An Empirical Analysis of the Federal Judiciary* 8–9 (Brookings 2006).

voting patterns, by individual judges, are found for Democratic appointees on panels consisting of only Democratic appointees, and when the most conservative patterns, by individual judges, are found for Republican appointees on all-Republican panels.[105]

2. The standard patterns here.

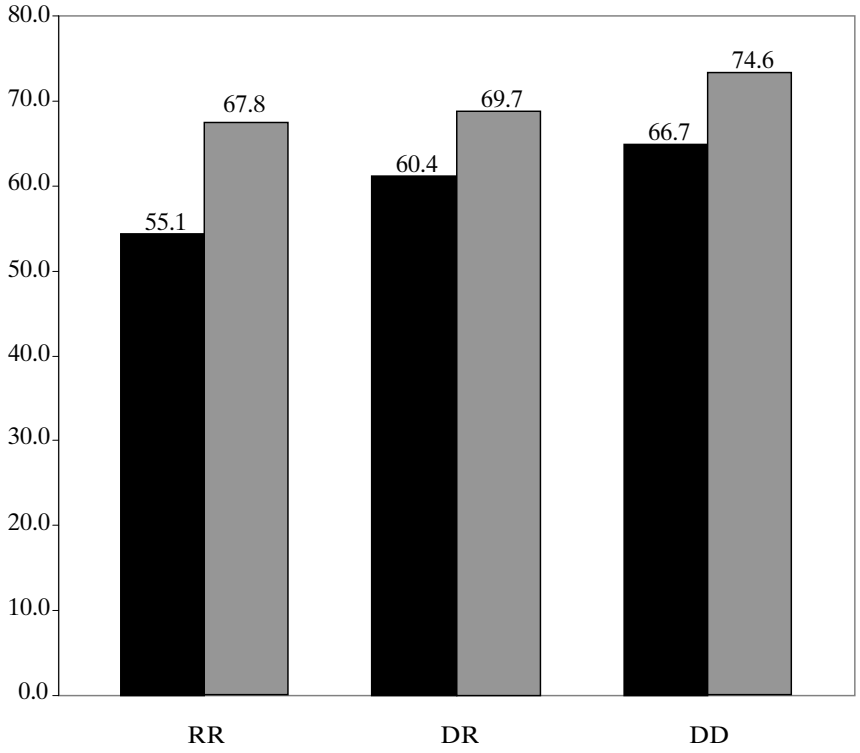
Democratic appointees typically show increasingly liberal voting patterns as the number of Democratic appointees increases, and Republican appointees typically show increasingly conservative voting patterns as the number of Republican appointees increases.[106] As we shall soon see, our most striking finding here is a form of ideological amplification, clearly demonstrated once agency and judicial decisions are coded in political terms. In arbitrariness cases, Democratic appointees show heightened liberal voting on all-Democratic (DDD) panels, just as Republican appointees show heightened conservative voting on all-Republican (RRR) panels.

Figure 2 examines whether panel effects are present in the distinctive context of validation rates. The rates of Democratic appointees appear in the lightly shaded bars and those of Republican appointees appear in darkly shaded bars. Panel effects are evident in the voting patterns of the appointees of presidents of both political parties. But they are more pronounced for Republican appointees. When sitting with two other Democratic appointees, the average Democratic appointee votes to validate 75 percent of the time. This rate falls by roughly 5 percentage points when the panel has one Republican appointee and by another 2 percentage points when it has two Republican appointees.

105 See, for example, Miles and Sunstein, 73 U Chi L Rev at 864 (cited in note 39).

106 See, for example, id at 860; Sunstein, et al, *Are Judges Political?* at 20–23 (cited in note 104).

FIGURE 2
VALIDATION RATES OF CIRCUIT COURT JUDGES
IN ARBITRARINESS REVIEW CASES BY PANEL COMPOSITION
AND BY PARTY OF APPOINTING PRESIDENT



Note: The darkly shaded bars indicate the validation rates of Republican appointees, and the lightly shaded bars indicate the validation rates of Democratic appointees.

Republican appointees demonstrate slightly greater responsiveness to panel composition. The validation rate of the average Republican appointee falls by a bit more than 6 percentage points when she sits with one Democratic appointee and one Republican appointee, rather than two Democratic appointees. When a panel consists of two other Republicans rather than a Democrat and another Republican, her average validation rate falls by 5 percentage points. We observe a modest form of ideological amplification.[107]

[107] This movement in the validation rates of Republican appointees implies a large difference in the validation rates of politically uniform panels. Even before considering the political direction of the agency decisions, the difference in the validation rates of all-Democratic and all-

As Table 1 reveals, overall validation rates obscure pronounced ideological patterns. Table 3 therefore decomposes the validation rates of Democratic and Republican appointees by both panel composition and the ideological content of the agency decision.[108] It compares the validation rates for judges appointed by presidents of each party according to whether the panel was politically "mixed" and whether the agency decision was liberal. Column (1) of Table 3 shows the overall validation rates, and these figures are comparable to those in Figure 2. But this grouping of the data showcases two points. First, the validation rates steadily decline as the number of Republican appointees on a panel grow. Second, the validation rates of Democratic and Republican appointees sitting on politically mixed panels are fairly close; they differ by 7 percentage points.

The remaining columns of Table 3 display the relationship of validation rates to the nature of the agency decisions. Two patterns are immediately evident. First, politically unified panels exhibit strong ideological responses to the content of the agency decisions. Too much importance should not be attached to the precise magnitudes of these differences because the number of votes in cases reviewing agency decisions is small. But the general patterns are clear. The average validation rate of a panel consisting of three Democratic appointees is 43 percentage points higher when the agency decision is liberal than when it is conservative! For panels consisting of three Republican appointees, the response is even stronger but in the opposite direction. The average validation rate of all-Republican panels is 29 percentage points lower when the agency decision is liberal than when it is conservative.

Republican panels is 17 percentage points. Again, this difference is due primarily to the drop in the validation rates of all-Republican panels.

[108] Because ideological content matters and the political affiliation of the president does not, we do not separately report panel effects according to that affiliation. We have analyzed the relevant data, however, and our findings are available on request.

TABLE 3
VALIDATION RATES OF CIRCUIT COURT JUDGES IN ARBITRARINESS REVIEW CASES BY PARTY OF APPOINTING PRESIDENT, BY IDEOLOGICAL CONTENT OF AGENCY DECISION, AND BY PANEL COMPOSITION

		Ideological content of agency decision			
	Panel composition	Total (1)	Liberal (2)	Not liberal (3)	Difference of (2)–(3):
Party of appointing president					
(A) Democrat	DDD	0.746 (0.037) [138]	0.812 (0.036) [117]	0.381 (0.109) [21]	0.431** (0.097)
(B) Democrat	DDR or RRD	0.689 (0.018) [669]	0.703 (0.019) [586]	0.590 (0.054) [83]	0.113** (0.054)
(C) Republican	DDR or RRD	0.619 (0.018) [767]	0.612 (0.019) [685]	0.683 (0.052) [82]	-0.071 (0.057)
(D) Republican	RRR	0.551 (0.025) [385]	0.526 (0.027) [352]	0.818 (0.068) [33]	-0.293** (0.090)
Difference of (A)–(B):		0.057 (0.043)	0.109** (0.045)	-0.209* (0.121)	
Difference of (A)–(C):		0.127** (0.044)	0.200*** (0.048)	-0.302** (0.116)	
Difference of (A)–(D):		0.196*** (0.048)	0.286*** (0.051)	-0.437*** (0.122)	
Difference of (B)–(C):		0.070** (0.025)	0.091*** (0.027)	-0.093 (0.075)	
Difference of (B)–(D):		0.138*** (0.030)	0.178*** (0.032)	-0.228** (0.096)	
Difference of (C)–(D):		0.069** (0.031)	0.086** (0.032)	-0.135 (0.092)	

Means, standard errors in parentheses, and number of observations in brackets.
Note: * denotes difference significant at 10 percent level, ** denotes difference significant at 5 percent level, and *** denotes difference significant at 1 percent level. Differences may not match exactly due to rounding.

These patterns imply that when the agency decision is liberal, the validation rate of a Democratic appointee sitting on a panel with two other Democratic appointees is 29 percentage points higher than that of a Republican appointee sitting with two other Republican appointees. When the agency decision is instead conservative, the direction of this gap reverses but its magnitude remains very large. The validation

rate of a Democratic appointee sitting with two other Democratic appointees is 44 percentage points below that of a Republican appointee sitting with two other Republican appointees. Here, then, is a clear "smoking gun" with respect to panel effects.

These figures reveal an important point: the seesawing validation rates of Democratic and Republican appointees in response to the nature of agency decisions (shown in Table 1) is largely attributable to the behavior of judges on politically unified panels. A comparison of Rows (A) and (D) in Table 3 shows a pattern of seesawing validation rates akin to that in Table 1—only more pronounced. The validation rates of Democratic appointees sitting with two other Democratic appointees are almost the mirror image of those of Republican appointees sitting with two other Republican appointees.

For judges sitting on politically mixed panels, the movement of validation rates in response to the ideological content of the agency decision is muted but not entirely absent. A Democratic appointee on a politically mixed panel has an average validation rate 11 percentage points higher when the agency decision is liberal rather than conservative, and this movement is statistically significant. A Republican appointee on a politically mixed panel votes to validate under arbitrariness review 7 percentage points less often when the agency decision is liberal rather than conservative, but this difference is not statistically significant.

These patterns also mean that when the agency decision is liberal, Democratic appointees on politically mixed panels vote to validate about 9 percentage points more often than do Republican appointees on politically mixed panels. The opposite happens when the agency decision is conservative; the average Democratic appointee on a politically mixed panel votes in favor of validation 9 percentage points less often than the average Republican appointee on a politically mixed panel. Even on politically mixed panels, Democratic and Republican appointees react to the ideological content of agency decisions in the predicted directions, and their responses are large enough to generate a seesaw pattern in validation rates. Notably, however, this pattern is less pronounced than on politically uniform panels.

3. Comparing *Chevron*.

As striking as these ideological patterns are, the role of judicial partisanship under *Chevron* is just as distinct, if not more so. In terms of raw numbers, the effect of such partisanship is even more dramatic under *Chevron*. The rate at which Democratic appointees sitting with two other Democratic appointees voted to validate liberal agency in-

terpretations of statutes was 32 percentage points higher than for conservative agency interpretations.[109] For Republican appointees sitting with two other Republican appointees, the validation rate was more than 40 percentage points higher when the agency interpretation was conservative than when it was liberal.[110] Hence it is plausible, but false, to speculate that *Chevron* has imposed a greater discipline on political voting than can be found in the domain of arbitrariness review. In our data, at least, the speculation is rejected.

When politically mixed panels reviewed agency interpretations under *Chevron*, the movement in validation rates of Republican appointees was not statistically significant, while for Democratic appointees it was an increase of 20 percentage points, which was statistically significant.[111] Interestingly, in arbitrariness review cases, Democratic appointees on mixed panels also show a statistically significant response to the nature of the agency decision while Republican appointees on mixed panels do not. Perhaps more importantly, the small sample sizes preclude any strong inferences about whether the response of politically unanimous panels to the nature of the agency decisions is larger under *Chevron* than under arbitrariness review. Nonetheless, the results in Table 3 indicate that when the ideological content of the agency decision is considered, the partisan composition of panels exerts a substantial influence on judges' exercise of arbitrariness review.

4. Conclusions.

The discussion of panel effects has been regrettably complex, but the major conclusions are plain. In arbitrariness review cases, judicial votes are significantly affected by the composition of the panel. The political party of the appointing president is a good predictor of judicial behavior in such cases, and the political party of the president who appointed the other two judges on the panel is also a strong predictor. A key finding is that the more Democratic appointees on a panel, the greater the likelihood that a panel will validate a liberal agency decision and the smaller the likelihood that it will validate a conservative one. But perhaps our most striking finding here involves the reaction of judges on politically uniform panels to the nature of the agency decisions. The willingness of judges on such panels to vote to validate an agency decision under the arbitrariness standard correlates strongly with the ideological direction of that decision. *On RRR and DDD panels,*

109 See Miles and Sunstein, 73 U Chi L Rev at 855–58 (cited in note 39).

110 See id.

111 See id.

judges are at least 20 percentage points more likely to reject an arbitrariness challenge when the agency decision fits with their presumed ideological preferences than when it does not. This finding suggests that the influences of judicial ideology and panel composition exert approximately the same power in arbitrariness review as they do in review under *Chevron*.

D. Liberal Voting, Conservative Voting

1. Liberal voting rates.

Another way to analyze the votes of the judges is to examine whether their votes can be considered "liberal." We classified a judge's vote as liberal if it was a vote either to validate a liberal agency decision or to invalidate a conservative agency decision. Table 4 presents comparisons of the liberal voting rates of Democratic and Republican appointees. Column (1) of Table 4 shows the overall liberal voting rates of Democratic and Republican appointees. Consistent with conventional wisdom, Democratic appointees cast liberal votes far more frequently. Democratic appointees vote in a liberal way 69 percent of the time, which is about 13 percentage points higher than the percentage for Republican appointees. It is worth underlining the fact that we are concerned with judicial decisions striking down agency judgments of fact or policy as arbitrary or unreasonable, where the convictions of federal judges are not supposed to play a role. But in that domain, a 13 percentage point difference shows a significant effect from judicial preferences.

This pattern is comparable to the liberal voting rates in cases reviewing EPA and NLRB decisions under *Chevron*. In such cases, Democratic appointees voted in a liberal way 67 percent of the time, and Republican appointees 50 percent of the time.[112] The slightly lower rate of liberal voting by Republican appointees in *Chevron* cases is not statistically distinguishable from their rate of liberal voting in arbitrariness cases. The partisan gap in liberal voting in cases reviewing agency decisions for arbitrariness is effectively identical to the gap in cases applying *Chevron*.

The next two columns of Table 4 break down the liberal voting rates according to the party of the current president. As seen previously, that party is a highly imprecise measure of the ideological content of the agency decisions in our data set. The final column of Table 4 shows that the liberal voting rates of both Democratic and Republican appointees rise by several percentage points when the party of the current president changes. For Democratic appointees, the direction of

---

[112] Id at 859.

this movement is contrary to the standard prediction. That is, Democratic appointees cast liberal votes slightly less often during Democratic Presidencies. But the difference is small, about 5 percentage points, and not statistically meaningful. The finding that Democratic appointees vote in a more liberal fashion during the Republican Presidency is consistent with our earlier claim that many of the NLRB decisions during Republican administrations are relevantly liberal.

TABLE 4
LIBERAL VOTING RATES OF CIRCUIT COURT JUDGES IN ARBITRARINESS REVIEW CASES BY PARTY OF APPOINTING PRESIDENT AND BY PARTY OF CURRENT PRESIDENT

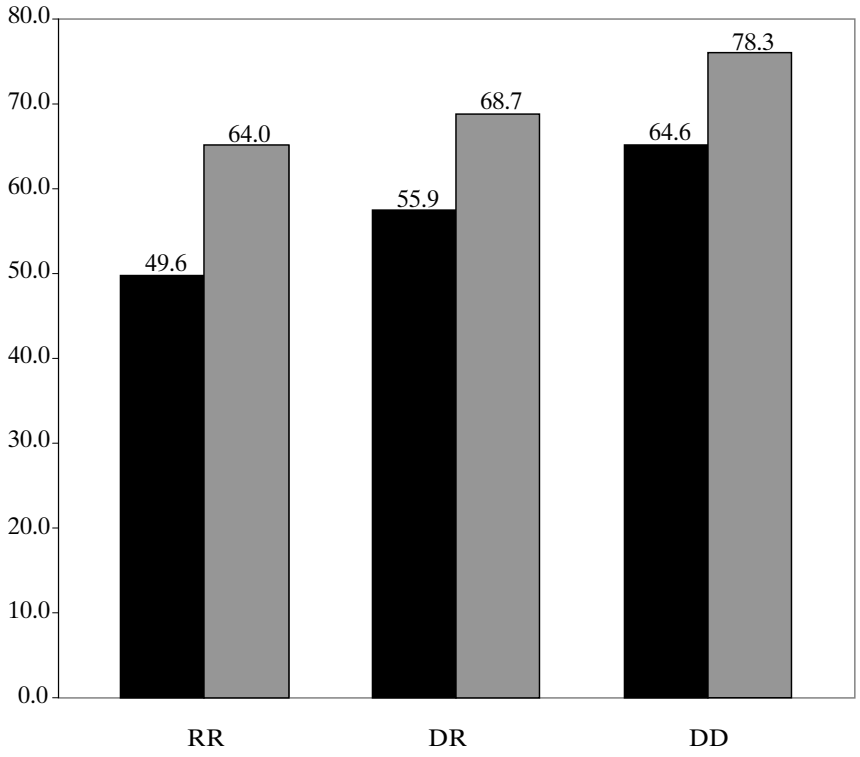
	Party of current president			
	Total (1)	Democrat (2)	Republican (3)	Difference of (2)–(3)
Party of appointing president				
(A) Democrat	0.686 (0.016) [807]	0.665 (0.023) [427]	0.711 (0.023) [380]	-0.045 (0.033)
(B) Republican	0.552 (0.015) [1,152]	0.520 (0.020) [617]	0.589 (0.021) [535]	-0.069** (0.029)
Difference of (A)–(B):	0.134*** (0.022)	0.145*** (0.031)	0.122*** (0.032)	

Means, standard errors in parentheses, and number of observations in brackets.
Note: * denotes difference significant at 10 percent level, ** denotes difference significant at 5 percent level, and *** denotes difference significant at 1 percent level. Differences may not match exactly due to rounding.

2. Panel effects.

Figure 3 presents liberal voting rates of the circuit judges by their political party and by the partisan composition of the panel. The liberal voting rates of Republican appointees are almost everywhere below those of Democratic appointees, but they are highest when the panel has a DDR configuration. In that setting, Republican appointees cast liberal votes about 65 percent of the time, which is 9 percentage points higher than when the panel has two Republican appointees and one Democratic appointee. Moreover, Republican appointees cast liberal votes most infrequently, 50 percent of the time, when the panel consists of three Republicans. Both ideological dampening and amplification thus characterize the liberal voting rates of Republican appointees.

FIGURE 3
LIBERAL VOTING RATES OF CIRCUIT COURT JUDGES
IN ARBITRARINESS REVIEW CASES BY PANEL COMPOSITION
AND BY PARTY OF APPOINTING PRESIDENT



Note: The darkly shaded bars indicate the validation rates of Republican appointees, and the lightly shaded bars indicate the validation rates of Democratic appointees.

Democratic appointees also show fluctuation in their liberal voting rates. When a panel consists of three Democratic appointees, they cast liberal votes 78 percent of the time—a rate that is 9 percentage points higher than when the panel has two Democratic appointees and one Republican appointee. Their liberal voting rates slip to 64 percent when the Democratic appointee sits with two Republican appointees. These movements are broadly consistent with familiar patterns in appellate decisionmaking. Compared to how a judge votes when sitting with one Democratic appointee and one Republican appointee, a judge sitting with two Democratic appointees is more likely to vote in a liberal way, and a judge sitting with two Republican appointees is less likely to vote in a liberal way. These patterns are

known as ideological amplification, meaning "amplified" ideological voting on RRR and DDD panels, and ideological dampening, meaning "dampened" ideological voting by Ds on DRR panels and by Rs on RRR panels. Ideological amplification and ideological dampening are evident in many areas of law, just as they are in the arbitrariness review cases depicted in Figure 3.[113]

Interestingly, the amplification effect for Democratic appointees appears slightly more pronounced than the dampening effect, while the opposite appears true for Republican appointees. We do not have an explanation for these intriguing differences in magnitude. Rather, the primary lesson of Figure 3 is that ideological amplification and dampening characterize the voting patterns of both sets of judges in arbitrariness review cases.

These results are in many ways consistent with our findings for *Chevron* cases, but some subtle differences are also present. Under both standards of review, Democratic appointees cast liberal votes more often than Republican appointees irrespective of the panel composition, and the liberal voting rates of judges appointed by presidents of both parties fluctuates to some degree with panel composition. But in the *Chevron* cases, the liberal voting rates of Democratic appointees, rather than Republican appointees, are more responsive to panel composition.[114] The liberal voting rates of Democratic appointees in *Chevron* cases climb steadily as the number of Democratic appointees on a panel grow.[115] The liberal voting rates of Republican appointees under *Chevron* are steady as long as Republican appointees form a majority of judges on the panel, and they dip when a Republican appointee sat with two Democratic appointees. While these patterns invite speculations about the differences between the two standards of review, the noisiness of the estimates prevents us from drawing strong inferences.

III. THE DISTINCTIVENESS OF THE COURT OF APPEALS FOR THE DISTRICT OF COLUMBIA

In view of its importance and its specialized docket, which consists in large part of regulatory problems, the Court of Appeals for the District of Columbia is the appellate court most frequently studied by administrative law scholars and political scientists.[116] Precisely because

113 See Sunstein, et al, *Are Judges Political?* at 54–56 (cited in note 104).

114 See Miles and Sunstein, 73 U Chi L Rev at 860–61 (cited in note 39).

115 See id.

116 See generally, for example, Revesz, 38 Va L Rev 1717 (cited in note 34); Richard J. Pierce, Jr., *Two Problems in Administrative Law: Political Polarity on the District of Columbia*

of its distinctive role, a natural question is whether the voting behavior of DC Circuit judges is representative. To what degree does the unique nature of the DC Circuit lead it to perform in distinctive ways? This Part provides some answers. Our most important findings are that the validation rate on the DC Circuit is significantly lower than the validation rate elsewhere, and that the court as a whole shows conservative voting patterns—so much so that Democratic appointees on the DC Circuit show voting patterns akin to those of Republican appointees elsewhere.

A. Three Initial Findings

Table 5 reports the rates at which Democratic and Republican appointees to the District of Columbia and other circuits vote to validate decisions of the EPA and the NLRB. Three aspects of these validation rates are immediately apparent. First, despite its specialized docket, the DC Circuit accounts for only 187 of the 653 cases, or somewhat less than a third of the total. This is a significantly lower share than in judicial applications of *Chevron*. Of the 227 challenges to the EPA and the NLRB under *Chevron* over the same time period, the DC Circuit decided 109 of them, or 48 percent.

Second, the final row of Table 5 shows that in both the DC Circuit and other circuits, the overall validation rates of Democratic appointees are higher than those of Republican appointees. In the DC Circuit, Democratic appointees vote to validate at a rate about 8 percentage points higher than that of their Republican colleagues, while in other circuits, they do so at a rate about 11 percentage points higher. The pervasiveness of this partisan gap is meaningful, but the intercircuit difference of 3 percentage points (that is, 8 versus 11) is not. Democratic appointees both within and outside of the DC Circuit voted to validate agency decisions significantly more often than Republican appointees.

The third finding is the most interesting. The final column of Table 5 shows that both Democratic and Republican appointees to the DC Circuit are much less willing to validate the decisions of the EPA and the NLRB in arbitrariness cases than are judges in other circuits. The Democratic appointees to the DC Circuit were 9 percentage points less likely to validate than their counterparts in other circuits, while for Republicans the difference was slightly smaller, 6 percentage points. This contrast is striking both for its size and for its consistency across partisan affiliations.

Circuit and Judicial Deterrence of Agency Rulemaking, 1988 Duke L J 300; Antonin Scalia, Vermont Yankee*: The APA, the D.C. Circuit, and the Supreme Court*, 1978 S Ct Rev 345.

TABLE 5
VALIDATION RATES OF CIRCUIT COURT JUDGES IN ARBITRARINESS REVIEW CASES BY PARTY OF APPOINTING PRESIDENT AND BY CIRCUIT

	Circuit court of appeals		
	District of Columbia (1)	All other circuits (2)	Difference of (1)–(2)
Party of appointing president			
(A) Democrat	0.634 (0.032) [232]	0.725 (0.018) [575]	-0.092** (0.036)
(B) Republican	0.552 (0.028) [328]	0.614 (0.017) [824]	-0.067* (0.032)
Difference of (A)–(B):	0.082* (0.042)	0.111*** (0.026)	

Means, standard errors in parentheses, and number of observations in brackets.
Note: * denotes difference significant at 10 percent level, ** denotes difference significant at 5 percent level, and *** denotes difference significant at 1 percent level. Differences may not match exactly due to rounding.

B. Of Familiarity and Contempt

A promising explanation of these differences is the greater experience of DC Circuit judges in reviewing administrative agencies. With the exception of a very few judges whose tenure on the bench only briefly overlapped with our sample period, every DC Circuit judge in this period appears at least two dozen times in our data.[117] The median number of votes in our data from such DC Circuit judges is 53, and the mean is 52.2. Outside of the DC Circuit, judges hear arbitrariness challenges to EPA and NLRB decisions far less frequently. The median number of judges deciding cases involving the EPA and the NLRB is 2, and the mean is 4.1.[118] If we were to calculate the me-

[117] Judges who left the DC Circuit relatively early in our observation period are Judges James Buckley, Patricia Wald, and then-Judge Ruth Bader Ginsburg, and those who joined it relatively late are Judges Janice Rogers Brown and Thomas Griffith. Then-Judge John Roberts is both a late arrival and an early departure in our data. The mean number of votes from each of these judges is 8.1 and the median is 6.

[118] Another example of this difference is the gap between the judges in our sample who have decided the most hard look reviews of the EPA and the NLRB. The DC Circuit judge in our data who voted in the most arbitrariness review cases, Judge Karen LeCraft Henderson, heard sixty-six of them. Outside of the DC Circuit, the judge in our data who sat on the most hard look cases, Judge Joel Flaum in the Seventh Circuit, heard twenty-five of them. Close behind Judge Flaum were his Seventh Circuit colleagues Judges Diane Wood and Ilana Rovner, each with twenty-four cases.

dian number of relevant EPA and NLRB cases heard by the typical appellate judge outside the DC Circuit, it would likely be zero because many judges never sit on cases requiring arbitrariness review and thus never appear in our data.

Judges who are experienced in conducting arbitrariness review might well become harder reviewers. With greater experience, judges may grow more confident in their own judgments about what is arbitrary and thus may be more willing to invalidate agency decisions. In addition to having greater experience with hard look review generally, judges may be more willing to invalidate if they have previously reviewed the decisions of a specific agency. Familiarity may not necessarily breed contempt, but repeated play should allow judges to learn where particular agencies are prone to weaknesses in their procedures or their analyses. It is reasonable to speculate that the higher invalidation rates in the DC Circuit, for both Democratic and Republican appointees, are at least partly explained in these terms.

C. Disaggregations

1. Validation rates.

Table 6 breaks down the comparison of the DC Circuit and other circuits by the ideological content of the agency decision. The standard seesaw pattern—in which the validation rates of Democratic appointees are higher than those of Republican appointees when the agency decision is liberal (and the opposite occurs when the agency decision is conservative)—is found only outside of the DC Circuit.

For DC Circuit judges, the validation rates of Republican appointees appear more responsive to the ideological content of the agency decision than those of their Democratic colleagues. Panel A shows that when the agency decision is liberal rather than conservative, the validation rates of Democratic appointees differ by only 4 percentage points. Moreover, this movement in the validation rates of Democratic appointees—higher when the agency decision is conservative—is contrary to the predicted direction. Too much should not be made of this slight increase because the number of cases in our data in which the DC Circuit reviewed conservative agency decision is small. The main conclusion is that the validation rates of Democratic appointees to the DC Circuit do not show the predictable ideological pattern. For Republican appointees on the DC Circuit, the difference in validation of conservative rather than liberal agency decisions is in the predicted direction and quite large—nearly 20 percentage points! On the DC Circuit, Republican appointees appear to show far more ideological voting than do Democratic appointees.

For judges on other circuits, the opposite is true. Panel B shows that when the agency decision is liberal, Democratic appointees validate at a very high rate, 76 percent, and when the agency decision is not liberal, their validation rate falls to 51 percent, a drop of 25 percentage points. The nature of the agency decision also has a sizable impact on the validation rates of Republican appointees. Republican appointees on appellate courts other than the DC Circuit vote to validate liberal agency decisions under hard look review about 60 percent of the time and conservative decisions about 72 percent of the time. The difference for Republican appointees outside the DC Circuit of 12 percentage points is about half as large as the 25 point difference for Democratic appointees—but it is still statistically significant.

In sum, the validation rates of Democratic and Republican appointees both inside and outside of the DC Circuit appear to respond to the ideological direction of the agency decisions. These responses produce the now-familiar seesawing of validation rates. Democratic appointees vote to validate at higher rates than Republicans when the agency decision is liberal, and vice versa when the agency decision is conservative. A general conclusion is that the influence of judges' political commitments on arbitrariness review is not limited to any particular circuit court.

TABLE 6
VALIDATION RATES OF CIRCUIT COURT JUDGES IN ARBITRARINESS REVIEW CASES BY PARTY OF APPOINTING PRESIDENT, BY IDEOLOGICAL CONTENT OF AGENCY DECISION, AND BY CIRCUIT

Panel A
Circuit Court of Appeals for the District of Columbia

	Ideological content of agency decision		
	Liberal (1)	Not liberal (2)	Difference of (1)–(2)
Party of appointing president			
(A) Democrat	0.629 (0.034) [205]	0.667 (0.092) [27]	-0.037 (0.099)
(B) Republican	0.531 (0.029) [292]	0.722 (0.076) [36]	-0.191** (0.087)
Difference of (A)–(B):	0.098** (0.045)	-0.056 (0.119)	

Panel B
Other Circuit Courts of Appeals

	Ideological content of agency decision		
	Liberal (1)	Not liberal (2)	Difference of (1)–(2)
Party of appointing president			
(A) Democrat	0.759 (0.019) [498]	0.506 (0.057) [77]	0.253*** (0.054)
(B) Republican	0.603 (0.018) [745]	0.722 (0.051) [79]	-0.119** (0.058)
Difference of (A)–(B):	0.156*** (0.027)	-0.215** (0.076)	

Means, standard errors in parentheses, and number of observations in brackets.
Note: * denotes difference significant at 10 percent level, ** denotes difference significant at 5 percent level, and *** denotes difference significant at 1 percent level. Differences may not match exactly due to rounding.

2. Liberal and conservative voting.

A possible explanation for the lower validation rates of DC Circuit judges is that they are simply more conservative than their colleagues on other circuits. The relative lack of responsiveness of Democratic appointees on the DC Circuit to the ideological content of the agency decision—as shown in Table 5—also suggests that Democratic appointees to that court may be more conservative than Democratic appointees on other courts of appeals.

To test this hypothesis, Table 7 presents comparisons for liberal voting rates. The table reveals that in addition to having higher invalidation rates, DC Circuit judges—of both parties—are significantly more conservative than judges of other circuits in their voting patterns in arbitrariness cases. Republican appointees to the DC Circuit cast liberal votes less often than their counterparts in other circuits by about 7 percentage points. The difference for Democratic appointees is almost double that: the average Democratic appointee to the DC Circuit casts liberal votes in these arbitrariness review cases about 13 percentage points less often than do Democratic appointees in other circuits.

TABLE 7
LIBERAL VOTING RATES OF CIRCUIT COURT JUDGES IN ARBITRARINESS REVIEW CASES BY PARTY OF APPOINTING PRESIDENT AND BY CIRCUIT

	Circuit court of appeals		
	District of Columbia (1)	All other circuits (2)	Difference of (1)–(2)
Party of appointing president			
(A) Democrat	0.595 (0.032) [232]	0.723 (0.019) [575]	-0.129*** (0.034)
(B) Republican	0.503 (0.027) [328]	0.572 (0.017) [824]	-0.069** (0.032)
Difference of (A)–(B):	0.092** (0.043)	0.152** (0.026)	

Means, standard errors in parentheses, and number of observations in brackets.
Note: * denotes difference significant at 10 percent level, ** denotes difference significant at 5 percent level, and *** denotes difference significant at 1 percent level. Differences may not match exactly due to rounding.

Yet the lower liberal voting rates of DC Circuit judges do not imply that Democratic and Republican appointees on the DC Circuit are equally conservative. In both the DC Circuit and other circuits, Democratic appointees vote in the liberal manner more often than Re-

publican appointees. This partisan gap is about 9 percentage points in the DC Circuit and 15 percentage points in other circuits, but the 6 point difference between these two figures is not statistically significant. We cannot conclude that the partisan gap is larger on other circuits, but we readily infer that Democratic appointees on all appellate courts cast liberal votes more often in arbitrariness review of EPA and NLRB decisions than Republican appointees do. We will explore shortly a complexity with drawing some tempting conclusions from this finding; for the moment, let us continue with the numbers.

While a partisan gap remains within each circuit, the lower liberal voting rates of Democratic appointees on the DC Circuit can be further illustrated by comparing them to Republican appointees in other circuits. Put differently, does the average Democratic appointee to the DC Circuit have liberal voting rates as low as the average Republican appointee to another circuit court? Table 7 shows, strikingly, that the answer is "yes." On the DC Circuit, Democratic appointees cast liberal votes 60 percent of the time, while Republicans on other circuits did so 57 percent of the time. The 3 percentage point difference between these figures is not statistically different from zero. At least in terms of this set of arbitrariness review opinions, DC Circuit Democrats behave like Republicans on other federal appellate courts. A look back at Table 5 reveals that these two groups of judges are also indistinguishable in terms of their validation rates in these cases. On the DC Circuit, Democratic appointees voted to validate the agency decisions 63 percent of the time, while Republican appointees on other circuits did so 61 percent of the time.

We have referred to a complexity in interpreting these findings, and it is easy to describe: arbitrariness cases in the DC Circuit might be relevantly different from arbitrariness cases in other courts of appeals. In some cases, the DC Circuit has exclusive jurisdiction to review EPA action, and pragmatic constraints might ensure that "liberal" challenges to agency action are relatively weak and that "conservative" challenges to agency action are relatively strong. In other cases, litigants might decide to bring particular cases in the DC Circuit, or decide not to do so, and these selection effects might defeat easy comparisons. We have therefore spoken of differences in liberal voting rates without drawing strong conclusions about whether DC Circuit judges are more conservative in the abstract. Because the mix of cases in the DC Circuit is not a random sample, our evidence is merely suggestive.

Nonetheless, the central lessons are plain. In cases challenging EPA and NLRB decisions for arbitrariness, judges on the DC Circuit vote to invalidate agency decisions more readily than do their counterparts on other circuits. Both Democratic and Republican appointees to the DC Circuit show lower liberal voting rates than do their

counterparts on other circuits. At the same time, the partisan gap between Democratic and Republican appointees does not lessen: in both the DC and other circuits, Democratic appointees cast liberal votes in arbitrariness review cases significantly more often than Republican appointees. In terms of their liberal voting rates, Democratic appointees to the DC Circuit are equivalent to Republican appointees to other courts of appeals.

IV. NORMATIVE ISSUES

We now turn to normative issues. It is tempting to think that an understanding of validation rates and of the role of judicial ideology would bear on and perhaps even resolve the continuing debate over hard look review. And in the end, we conclude that our findings can reasonably be taken to suggest the importance of diminishing that role and also provide significant ammunition for those who believe that such review should be softened.[119] But this lesson is heavily qualified, and the qualifications may be as important as the ultimate conclusion. The less ambiguous lesson is that it would be highly desirable to reduce the role of judicial policy preferences in conducting arbitrariness review.

A. Problems and Puzzles

Let us begin with the validation rate. In arbitrariness cases, the validation rate is 64 percent.[120] Some people might find that rate alarmingly low. After all, agencies are supposed to lose not when they are wrong but when their judgments of policy and fact are "arbitrary" (or lacking substantial evidence). An invalidation rate of 36 percent seems quite high.

To make progress on the normative issues, suppose we found that the validation rate was lower than it actually is—say, 30 percent. Would it be appropriate to conclude that the hard look doctrine was too hard? At first glance, the answer would clearly be affirmative. If courts are striking down agency decisions as arbitrary more than half of the time, there is reason to suspect that something is seriously amiss.

But for two reasons, the suspicion might turn out to be wrong. Perhaps more than half of the agency decisions are, in fact, arbitrary. If

---

[119] See, for example, Pierce, 47 Admin L Rev at 65 (cited in note 30) (arguing that court interpretations have transformed the efficient process envisioned by the APA into an overly inefficient and ineffectual process).

[120] We have explored NLRB and EPA cases here, but we have also compiled a different data set involving all decisions citing *State Farm*. When including all such cases, the validation rate is under 60 percent. See generally Miles and Sunstein, *The Hard Look in Practice* (cited in note 76).

so, the invalidation rate is nothing to deplore. But there is a more subtle point. As we have suggested,[121] the rate of challenges to agency action will be affected by whether judicial review is aggressive or weak. If courts are aggressive, we should expect to see more challenges, simply because the likelihood of success is higher. If the rate of challenges varies with the stringency of judicial review, then we might hypothesize that it would hover around a fairly constant level—as a first approximation, 50 percent (not so far from the overall rate that we in fact find). In other words, the rate of validation might be impervious to changes in the stringency of review.

To see why this is a plausible hypothesis, imagine that the stringency of judicial review was reduced, in the next five years, by about half—so that the validation rate would be 82 percent, all else equal, in five years. The difficulty is that the mix of cases would be most unlikely to remain constant. If litigants are rational, the likelihood of success will affect their decision whether to litigate, and that likelihood will depend on the aggressiveness of arbitrariness review. In all probability, many challenges that would have been brought would no longer be brought, simply because such challenges would be a waste of time and money. Even if the stringency of judicial review was cut in half, the overall validation rate could remain 64 percent. On certain assumptions about litigant behavior, less stringent review could even produce a *lower* validation rate, if, for example, only very strong cases were brought. (Hence the relatively higher validation rate for the EPA than for the NLRB may tell us less than first appears.)

Actually things are more complicated still. Agency decisions should also be affected by the likelihood of judicial invalidation. Consider the extreme case of no review, at all, of agency judgments about policy and fact. Without any such review, some agencies would inevitably make some decisions that they will not now make; the rate of arbitrariness would significantly increase. Of course it is likely that arbitrary decisions are already checked by nonjudicial safeguards of various sorts,[122] and we could imagine a world in which the level of arbitrariness would be very low even without judicial review. But in our world, it is more than reasonable to think that judicial review operates to discourage some decisions, actually or arguably arbitrary, that would be made in its absence.[123] If this is true, then it is also more than

---

121 See note 93 and accompanying text.

122 Consider, for example, the process of internal executive branch review. See Breyer, et al, *Administrative Law* at 102–13 (cited in note 27) (describing presidential checks on agency policy through executive orders establishing presidential oversight groups).

123 See Pederson, 85 Yale L J at 59–60 (cited in note 5) (contending that aggressive judicial review has helped to discipline arbitrary decisions by the EPA).

reasonable to think that aggressive review will discourage more decisions than weak review. And if this is true, then aggressive review will operate as a check on its own use. With such review, the mix of agency decisions will shift in the direction of less arbitrariness, and hence the rate of invalidation might well stay constant.

The analysis is analogous to that of the selection of disputes for litigation. In the standard account, the only disputes that advance to trial are those in which each party is sufficiently optimistic about her chance for success at trial that her estimated return from trial exceeds the difference between trial costs and settlement costs. Less optimistic litigants will choose to settle.[124] Arbitrariness review differs in many ways from trials, but the relevant insight is that both the rate of challenges to agency decisions and the content of agency decisions will respond to the intensity of judicial review.

We can therefore identify two ways in which changes in the intensity of judicial review will influence the case mix: as review becomes less intense, litigants will challenge fewer decisions and agencies will be more likely to make decisions that aggressive courts would have struck down as arbitrary. In terms of validation rates, the two effects will cut in different directions. Fewer challenges will mean lower validation rates; agency adaptation, in the form of decisions closer to the line of arbitrariness, will mean higher ones.

To know the ultimate consequences of less intense review, we need to know not only the direction but also the magnitude of these two effects. A reduction in the intensity of review should first tend to raise the validation rate. If neither litigants nor agencies were responsive to the intensity of review, an increase in the validation rate would be the sole consequence of less intense review. But if litigants were highly responsive to intensity shifts and if agencies were not, the content of agency decisions would remain the same while litigants would decline to bring the more marginal challenges. The volume of arbitrariness challenges would decline, and the validation rate could remain fairly constant. Or the validation rate could even fall as litigants found it worthwhile to challenge only agency decisions that were most egregiously wrong and thus most likely to be invalidated.

---

[124] Priest and Klein, 13 J Legal Stud at 12–13 (cited in note 93). A strong version of the Priest-Klein hypothesis predicts that under certain circumstances the win rate of plaintiffs at trial approaches 50 percent. We do not consider the fact that the validation rate in our EPA and NLRB cases hovers near 50 percent to be evidence for this proposition because the conditions for the Priest-Klein result are not satisfied in *State Farm* cases. Most important, the stakes in judicial review of agency decisions—however the relevant figures might be defined—are unlikely to be symmetrical between plaintiffs and defendants. Just one of the several possible ways in which the stakes may be asymmetric is that the agency, as a policymaking institution, may have greater concern for the precedential effect of the litigation than an individual litigant.

Suppose, however, that litigants were not responsive to intensity shifts and agencies were highly responsive. If so, then agencies might issue more arbitrary decisions without suffering a reduction in the validation rate and without inducing an increased flow of challenges. But if the agency let the quality of its decisions decline too much, the greater arbitrariness of decisions would eventually offset the reduction in the intensity of judicial review. In this instance, the volume of challenges might rise and the validation rate might fall. If both litigants and agencies were highly responsive to intensity shifts, the impact on the quality of agency decisions, the volume of challenges, and the validation rate would be far from clear. Without strong assumptions about whether litigants or agencies are more responsive to the intensity of judicial review, the overall impact on the validation rate is hard to foresee. These predictions are captured in Table 8.

It follows that in the abstract, it is difficult to draw firm conclusions about the stringency of review, and about whether existing practice is too lenient or too stringent, from any particular validation rate. In the future, much progress might be made by examining the rate of challenges to decisions by particular agencies or agencies in general—an eminently feasible enterprise. What percentage of EPA rules is actually challenged? Is the percentage higher now than it was five years ago, or ten years ago, or fifteen years ago? Still more progress might be made by a qualitative assessment of the nature and rationality of agency decisions over time—also feasible even if more difficult.

TABLE 8
EFFECTS OF REDUCED INTENSITY OF JUDICIAL REVIEW ON VALIDATION RATES

	Is the agency responsive?	
Are the litigants responsive?	Yes	No
Yes	Ambiguous impact on validation rate	No change (or lower validation rate)
No	No change (or lower validation rate)	Higher validation rate

B. Republican Appointees, Democratic Appointees, and the Smoking Gun

At this point, the validation rates that we observe might be taken to be insufficiently informative about whether courts are reviewing agency decisions too aggressively. But we are particularly interested in party and panel effects, and we should be able to learn more if we attend to the differences between Republican and Democratic appointees.

To focus the analysis, suppose that Republican appointees voted to invalidate liberal agency decisions 90 percent of the time and to invalidate conservative agency decisions 10 percent of the time—and that Democratic appointees showed the opposite pattern. Or suppose that Republican appointees showed an 80 percent validation rate during Republican administrations and a 20 percent validation rate during Democratic administrations—and that Democratic appointees showed a similar form of favoritism. At first glance, voting patterns of this kind would suggest a serious problem in the real world of arbitrariness review.

If these were the observed patterns, then we might be tempted to say that when an agency decision is invalidated as arbitrary, it is not always because it is genuinely arbitrary. On the contrary, it may well be because the court would have preferred the agency to do otherwise. At least this is so when Republican appointees vote to strike down liberal decisions or those in Democratic administrations, and when Democratic appointees vote to strike down conservative decisions or those in Republican administrations.

Even here, however, it is necessary to be careful. Suppose that Republican appointees strike down liberal agency decisions at a much higher rate than conservative agency decisions. By itself, does this fact *demonstrate* bias? The answer is that it does not. Perhaps liberal agency decisions are especially likely to be arbitrary. And if Democratic appointees are peculiarly likely to strike down EPA decisions under Republican presidents, it may be because such decisions are indeed arbitrary. After finding an asymmetrical pattern of votes within any particular group of appointees, we might well be suspicious of ideological bias on the part of the judiciary. But in truth, no such suspicion has been vindicated by that kind of pattern.

The smoking gun, we think, is the seesaw pattern found in Table 1—the fact that Republican validation rates jump from 58 percent to 72 percent when the agency decision becomes conservative, just as the Democratic validation rates fall from 72 percent to 55 percent when the agency decision becomes conservative. It cannot be the case that *both* camps are bias-free, simply responding to what any objective observer would deem arbitrary. The existence of ideological amplification sharpens this claim. If Democratic appointees show a greater rate of liberal voting when sitting with two Democratic appointees, and if Republican appointees show a greater rate of conservative voting when sitting with two Republican appointees, then something does seem seriously wrong.

To be sure, it is *possible* that one group is essentially neutral and right and that the other group is not. But it would be surprising if this possibility could be confirmed by an independent observer who was both neutral and right. The best conclusion is that in its operation, ar-

bitrariness review is significantly affected by the ideological dispositions of federal judges in a way that produces serious errors in light of the aspirations of *State Farm* itself. Recall that the most fundamental justification of hard look review is that with the grant of broad discretionary power to regulatory agencies, a firm judicial check is necessary as a kind of second-best substitute for insistence on the safeguards of the original constitutional system.[125] If the consequence of that firm check is to give effect to the policy commitments of federal judges, the cure seems worse than the disease.

Notice, however, that our data show a large but not massive difference between Republican and Democratic appointees. It is not as if Republican appointees have a 10 percent validation rate when the agency decision is liberal and a 90 percent validation rate when the agency decision is conservative. On the contrary, Republican appointees vote to validate most liberal agency decisions, and Democratic appointees vote to validate most conservative agency decisions. A defender of the status quo, seeking to minimize the role of judicial policy preferences, might respond with the suggestion that the evidence is compatible with the view that *State Farm* has disciplined the judicial role, ensuring as it has that Republican and Democratic appointees generally agree with one another.

We strongly agree with this suggestion insofar as it is meant to suggest that our data demonstrate that judicial ideology is not playing a dominant role and that judicial policy choices are not driving arbitrariness review. A crudely "realist" picture of existing practice is wildly inconsistent with reality. Nonetheless, judicial policy preferences do play a significant role, and in the difficult cases, it does seem to be driving actual outcomes. Something is seriously amiss if Republican appointees are significantly more likely to uphold conservative agency decisions than liberal agency decisions and if Democratic appointees show the opposite pattern. We cannot rule out the possibility that one group has it essentially right. But it is not possible that both groups have it essentially right, and we suspect that errors can be found from both sides.

C. Too Few Invalidations? Of Suspicion and Loyalty

At this stage, politically motivated invalidations might seem to be the most serious problem. But a different reading of our findings is imaginable. Perhaps the problem is not that appointees of both parties

---

[125] For discussion, see Sunstein, 1983 S Ct Rev at 181–84 (cited in note 20) (describing the procedural and substantive components of the hard look doctrine). See also text accompanying notes 7–9.

vote, with some regularity, to invalidate decisions with which they might be expected to be unsympathetic. Perhaps the real problem is that appointees of both parties vote, with some regularity, to *uphold* the decisions of agencies with which they might be expected to be sympathetic. Perhaps the real problem, uncovered by our data, is not politically driven suspicion but instead politically driven loyalty or at least sympathy. On this view, what is most troubling, and what emerges as the real story here, is the high rates of validation, by judges nominated by a president of one or another party, of agency decisions that those judges might be expected to find agreeable.

Nothing in our data excludes this possibility. If the challenged agency decisions are often arbitrary, perhaps it is disturbing to see that Democratic appointees vote to uphold liberal decisions and that Republican appointees vote to uphold conservative decisions. Perhaps it is affirmatively desirable to find a high level of invalidations; perhaps the correct rulings are those by Republican appointees of liberal decisions and those by Democratic appointees of conservative decisions. In a world in which agency arbitrariness is pervasive, politically driven validations would indeed be the problem.

Even if this is the correct reading of the data, and if validations rather than invalidations are the problem, the existing pattern of outcomes cannot be defended. If the real world of arbitrariness review includes a significant degree of political voting, leading to an unduly high validation rate, the appropriate correction is a "harder look," in the form of a general increase in judicial scrutiny. Moreover, we suspect that this is not the appropriate correction. If Democratic appointees are striking down conservative decisions at a higher rate than liberal decisions, and if Republican appointees are doing the same with liberal decisions, it would be surprising to find that all or most of the invalidated decisions are genuinely "arbitrary" within ordinary understandings of that term while also finding that the validated decisions are genuinely not "arbitrary." Most of the cases in our data involve complex questions of fact and policy on which reasonable people can differ. But we do not deny the potential value of a more systematic inquiry into the possibility that politically driven validations are a real problem.

D. What Should Be Done? Of Decision Costs and Error Costs

How do our findings bear on the continuing debate over arbitrariness review?[126] The first point is that questions about such review

126 See, for example, Pierce, 47 Admin L Rev at 65 (cited in note 30) (arguing that "courts have transformed the simple, efficient notice and comment process into an extraordinarily

cannot be settled in the abstract. Imagine, for example, a parallel world in which agency decisions are almost never arbitrary and never especially harmful even when arbitrary. Imagine that in such a world, judicial review of arbitrariness would produce more, not less, in the way of arbitrariness, simply because judicial decisions are replete with bias and suffer from a lack of expertise and accountability. In that world, there is no point to arbitrariness review. By hypothesis, such review would make the situation worse rather than better. These are claims about the costs of errors; perhaps arbitrariness review increases those costs. At the same time, judicial review increases the costs of decisions, simply because it adds an additional layer, and possibly more than that, of decisional burdens on all sides.

Some people, in some periods, have believed that the United States is not so far from this imaginary world.[127] At the very least, it is possible to find periods in which prominent voices suggest that aggressive judicial review of agency judgments of policy and fact is likely to cause serious problems.[128] The New Deal period is a prominent example;[129] the same is true of the period after the election of President Reagan.

At the same time, we could easily imagine another and quite different parallel world, in which agency decisions are distinctly susceptible to the influences of self-interested private groups, or otherwise a product of bias and confusion. In that world, we might also suppose that federal courts would provide an important ex post corrective and ex ante deterrent to biased and confused decisions.[130] In such a world, stringent judicial review of agency judgments of fact and policy would be easy to defend.

---

lengthy, complicated, and expensive process"); Mashaw and Harfst, *Auto Safety* at 95 (cited in note 10) (asserting that hard look review prevented the NHTSA from implementing its regulatory plan); Breyer, 38 Admin L Rev at 363 (cited in note 10) (advocating analysis and reform of doctrine requiring judicial deference to an agency's interpretation of a statute but more thorough review of an agency's factual and policy determinations).

127 See, for example, Mashaw and Harfst, *Auto Safety* at 249 (cited in note 10) (contending that the courts' tendency to "proceduralize" their judicial review of agency rulemaking has "imposed a debilitating, defensive posture on agency standard setting").

128 See, for example, id at 95 (describing how the NHTSA largely "abandoned" its new regulatory plan after negative outcomes in court cases); James M. Landis, *The Administrative Process* 154–55 (Yale 1938) (arguing that, although judges are "experts in the synthesis of design," they are not experts in other modern disciplines and so should apply deferential review to agency policymaking done by experts).

129 See Breyer, et al, *Administrative Law* at 19–20 (cited in note 27) (describing New Deal–era arguments that judicial decisions limiting regulatory power "were obstructing the popular will and thwarting the economic survival of the nation").

130 See Pederson, 85 Yale L J at 59–60 (cited in note 5) (claiming that internal review is not as effective as aggressive judicial review).

We can identify prominent voices, in prominent periods, suggesting that this understanding is not so far from reality.[131] The enactment of the APA was based on concerns about agency bias and relative enthusiasm for judicial review.[132] Indeed, the post–New Deal strengthening of substantial evidence review resulted from the Supreme Court's recognition of Congress's expression of a "mood" in favor of a more aggressive approach from the courts.[133] The rise of the hard look doctrine in the 1960s and 1970s was founded on similar assumptions about the value of judicial review in correcting agency errors and bias.[134] Fearful of agency "capture," prominent judges defended the doctrine as a method for increasing agency accountability by ensuring attention to the claims of a range of relevant interests, and also as a method of promoting the application of technical expertise to difficult problems.[135] For their part, skeptics pointed to the risk that hard look review would discourage agency rulemaking and reflect judicial bias on the merits.[136]

Our own findings demonstrate that judicial commitments are playing a significant role—and suggest the strong possibility that in many cases, judges are voting to invalidate agency decisions as arbitrary when they would not do so if their own predilections were otherwise. To the extent that this is so, there is a new argument for a softer look—that is, one that would ensure that agency decisions would be invalidated as arbitrary only when diverse judges could agree that they should be invalidated for that reason. We have a degree of confi-

131 See, for example, Leventhal, 122 U Pa L Rev at 536–37 (cited in note 2) (worrying that agencies will create policies by highly informal means without adequately synthesizing the record if courts do not prod them).

132 See Breyer, et al, *Administrative Law* at 20–21 (cited in note 27) (describing the passage of the APA as a compromise addressing "the largest perceived problem: administrative discretion"); *Universal Camera*, 340 US at 477–87 (highlighting the congressional history and widespread dissatisfaction with the abilities of courts to review procedures used by administrative bodies).

133 See *Universal Camera*, 340 US at 487 (asserting that this "mood" is evident in both the legislative history and the legislation itself).

134 See Sunstein, 1983 S Ct Rev at 177 (cited in note 20) (contending that as a result of the New Deal and the post–World War II growth of the "vast administrative apparatus," a need for agency discipline emerged beyond the inadequate political supervision in place).

135 See, for example, Leventhal, 122 U Pa L Rev at 555 (cited in note 2) (arguing that judicial review of agency decisions adds a degree of "effective supervision with restraint" and ensures that agencies take congressional mandates into account). Compare *Ethyl Corp v EPA*, 541 F2d 1, 67 (DC Cir 1976) (Bazelon concurring) (advocating that judicial attention should be focused on procedural rather than substantive matters), with id at 68–69 (Leventhal concurring) (arguing that judges cannot abstain from substantive review, even in technical matters, because courts have a congressionally mandated responsibility to ensure that agency decisionmaking is not irrational or discriminatory).

136 See, for example, Mashaw and Harfst, *Auto Safety* at 149–51 (cited in note 10); Melnick, *Regulation and the Courts* at 11 (cited in note 31) (asserting that judges "have become increasingly willing to second-guess agencies . . . especially when they fear the agency lacks aggressiveness in pursuing its statutory mission").

dence in the result when conservative decisions are invalidated by panels of RRD and RRR; we also have such confidence when liberal decisions are invalidated by panels of DDR and DDD. The troublesome cases involve invalidations that "fit" with the presumed ideology of the majority of judges on the relevant panel.[137] And indeed, we have found disturbing patterns of that kind on the courts of appeals. But as we have acknowledged, it is also possible to be troubled by validations that fit with the presumed ideology of that same majority of judges. We might add that we have a degree of confidence in validations from DDR and DRR panels, and from DDD and RRR panels, that are inconsistent with political expectations—but that there is room for concern when an RRR panel upholds a conservative decision and when a DDD panel upholds a liberal one.

It would therefore be possible to take our findings to support two different positions. The first is that judicial review should generally be weakened, so as to diminish the risk that invalidations reflect political commitments on the part of the relevant judges. The second is that steps should be taken to reduce the risks associated with potentially partisan validations or invalidations—as, for example, when DDD panels regularly uphold liberal agency decisions and when RRR panels show a special willingness to strike down such decisions. The second position seems to us more securely grounded in the evidence. It is certainly possible to fear that a general softening of judicial review would reduce a valuable ex post corrective and ex ante deterrent to arbitrary decisions—a fear to which our data do not speak.

Our findings also generate a clear prediction: when a judiciary dominated by the appointees of a Republican president reviews liberal agency decisions, or when a judiciary dominated by Democratic appointees reviews conservative agency decisions, the invalidation rate will increase. This prediction suggests that the debate over *State Farm* should consider the temporal effects of judicial policy preferences. The life tenure of federal judges implies that partisan imbalances in the appellate courts may persist for long periods. An unbalanced federal judiciary might well act as a brake on agencies' ability to implement the liberal or conservative policies of a new executive.[138]

---

137 Consider Frank B. Cross and Emerson H. Tiller, *Judicial Partisanship and Obedience to Legal Doctrine: Whistleblowing on the Federal Courts of Appeals*, 107 Yale L J 2155, 2172 (1998) (finding that unified panels show particular partisanship in the predicted ideological direction in a sample of *Chevron* cases).

138 Consider Adam B. Cox and Thomas J. Miles, *Judging the Voting Rights Act*, 108 Colum L Rev 1, 49–50 (2008) (arguing, in the context of the preclearance procedures of the Voting Rights Act, that partisan influence in the DOJ may be more variable over time than judicial partisanship).

The intensity of arbitrariness review can conspire with the life tenure of judges to make the effects of judicial ideology enduring.

It is not clear whether large-scale reforms are desirable. But several lessons do seem plain. At a minimum, the argument for Supreme Court review is strengthened in arbitrariness cases in which the outcome fits with the predicted ideological dispositions of unified panels. So too, those circumstances present unusually strong arguments for en banc review. A parallel lesson is more subtle and perhaps more important. If appellate judges are made aware that the evidence suggests a degree of ideological voting in arbitrariness review, perhaps that very awareness can operate as a kind of corrective or inoculation. In fact, our hope is that lawyers' and judges' understanding of the data might help to reduce the relevant effects in the future.

More generally, there might well be a fresh reason to revisit the current hard look review as a means of reducing the risk that agency decisions will be deemed arbitrary simply because judges do not agree with them on the merits. Remarkably, the Supreme Court has issued no major pronouncements about arbitrariness review since *State Farm* itself. Its next encounter with the topic could provide a context for directing significant cautionary notes to the courts of appeals.

We could also imagine more dramatic responses. Here, as elsewhere, there is reason to prefer mixed to unified panels, as a way of reducing the risk of ideologically driven outcomes.[139] In an important context, Congress has made exactly this choice, ensuring that the independent agencies may have no more than a bare majority of commissioners from the same political party.[140] For example, the NLRB, the FTC, the FCC, and the SEC must have at least two Republican members (of five) under Democratic presidents, and at least two Democratic members under Republican presidents.[141] This requirement might well operate to reduce the risks of ideological outcomes that would arise if adjudicative bodies consisted only of Republican or Democratic appointees. Note that the NLRB is one of our two principal agencies in the current study and that the NLRB makes almost all of its law and policy through adjudication, construing the provisions of

---

139 James Stribopoulos and Moin A. Yahya, *Does a Judge's Party of Appointment or Gender Matter to Case Outcomes?: An Empirical Study of the Court of Appeal for Ontario*, 45 Osgoode Hall L J 315, 362–63 (2007) (stressing the importance of diverse panels of judges with regard to party and gender); Sunstein, et al, *Are Judges Political?* at 135–36 (cited in note 104) (arguing that diverse panels are more likely to follow the law correctly).

140 See, for example, 15 USC § 78d(a) (2000) ("Not more than three of [the five SEC] commissioners shall be members of the same political party.").

141 See id.

the NLRA.[142] It may be worth considering steps that would ensure mixed panels on appellate courts, at least in high-stakes cases involving review of agency decisions for arbitrariness.

A more ambitious plan would enlist voting rules—by, for example, requiring unanimous decisions for invalidation of agency rulings as arbitrary.[143] We do not believe that our data support such a plan; the extent of ideological voting cannot justify such a dramatic departure from standard practices. But if unified panels are found, in the future, to show highly ideological voting patterns, it would be important to take steps to ensure that arbitrariness review does not amount, in practice, to Democratic or Republican review.

CONCLUSION

In *State Farm*, the Supreme Court attempted to establish a framework that would check arbitrariness on the part of administrators who are often given broad discretionary authority.[144] Our principal goal here has been to investigate the real world of arbitrariness review. We have found that the validation rate for the NLRB and the EPA is 64 percent. This rate is remarkably close to the validation rate for similar cases under *Chevron*.

The more important finding is that Democratic appointees show a far higher rate of liberal voting than do Republican appointees: 69 percent as opposed to 55 percent. When agency decisions are liberal, Democratic appointees are significantly more likely to vote to uphold them than when they are conservative. By contrast, Republican appointees are significantly more likely to uphold conservative agency decisions than liberal agency decisions. Democratic appointees show especially liberal voting patterns when sitting on all-Democratic panels; Republican appointees show especially conservative voting patterns when sitting on all-Republican panels.

It follows that the political party of the appointing president is a fairly good predictor of how a judge will vote in cases involving arbitrariness review; but the political party of the president who appointed

---

142 The Board's famous tendency to avoid rulemaking is noticed, with an evident lack of enthusiasm, in *Allentown Mack Sales & Service, Inc v NLRB*, 522 US 359, 374 (1998) (noting that the NLRB, "uniquely among major federal administrative agencies," has only adopted one regulation and otherwise promulgates rules entirely through adjudication).

143 See Jacob E. Gersen and Adrian Vermeule, Chevron *as a Voting Rule*, 116 Yale L J 676, 680 (2007) (arguing that requiring a unanimous panel decision would internalize deference and maintain a constant level of deference over time).

144 Note that the National Traffic and Motor Vehicle Safety Act asks the secretary of transportation to issue standards that "shall be practicable, shall meet the need for motor vehicle safety, and shall be stated in objective terms." 15 USC § 1392(a), repealed by Pub L No 103-272 § 7(b), 108 Stat 1379 (1994).

the two other judges on the panel is also a strong predictor. These conclusions might be taken to provide fresh support for those who seek to soften arbitrariness review, or at least for those who seek to reduce the role of judicial policy preferences in review of agency action. The hard look doctrine is most plausibly justified as a method for controlling the exercise of open-ended authority by regulatory agencies. To the extent that the doctrine operates, in practice, as a method of substituting judicial policy preferences for agency policy preferences, something is seriously wrong. Whether or not general softening is in order, steps might be taken to reduce the risk that judicial policy preferences are producing unjustified invalidations (and perhaps validations).

But our emphasis has been empirical, not normative. While the differences between Republican and Democratic appointees are significant, most judicial votes are not driven by political convictions. Recall that Republican appointees generally vote to validate liberal agency decisions and that Democratic appointees generally vote to validate conservative agency decisions. For this reason, it would be far too simple to say that the hard look doctrine is operationalized in purely political terms; our evidence is inconsistent with that conclusion. But it would not be too much to say that in important domains, the hard look is hardened, or softened, by the political predilections of federal judges.