

REASONABLE AGENCIES

*David Zaring**

THIS Article argues that the complex doctrine of judicial review of administrative action—containing no less than six separate tests depending on the sort of agency action to be reviewed—both descriptively is and normatively should be simplified into a “reasonable agency” standard. Courts, following step two of the Chevron doctrine, have started to sneak a reasonableness standard into their review in lieu of making the difficult distinctions required by current doctrine. Scholars evaluating the difference among the various doctrinal tests have started to note the increasing similarity among the tests, at least as applied by the courts. Empirical research, to which this Article contributes an additional study, suggests that regardless of the standard of review, courts affirm agencies’ actions slightly more than two thirds of the time; the variance of the validation rates of agency action, regardless of the standard of review, is small. A reasonable agency standard would simplify and clarify administrative law, better describe what courts actually do when confronted with agency action, and better explain the judicial role in the administrative state.

INTRODUCTION.....	136
I. STANDARD OF REVIEW DOCTRINE.....	143
A. Review of Law	143
B. Review of Fact.....	147
C. Arbitrariness Review	150
D. Implications of the Six Standards of Review.....	152
II. THE REALITY OF ADMINISTRATIVE REVIEW	153
A. Review of Law Is Becoming Review for Reasonableness	155

* Assistant Professor, Legal Studies Department, Wharton School of Business. Thanks to Amanda Frost, David Law, Amanda Leiter, Chad Oldfather, Brad Snyder, Jonathan Siegel, Robert Tsai, and Adrian Vermeule for helpful comments. Thanks also to Paul Fattaruso and Hang Xu for research assistance, and to Blake McShane for statistical advice.

136	<i>Virginia Law Review</i>	[Vol. 96:135
	<i>B. Arbitrariness Review Is Inseparable from the Reasonableness Inquiry that Is the Basis for Review of Law</i>	162
	<i>C. Arbitrariness Review and Fact Review Are Identical Standards</i>	166
III.	STANDARDS OF REVIEW IN PRACTICE.....	169
	<i>A. Prior Scholarship on Agency Validation</i>	170
	<i>B. Judicial Review of Agency Factfinding</i>	177
CONCLUSION	186
	<i>A. The Case for Current Doctrine and Its Discontents</i>	187
	1. <i>Clarity</i>	187
	2. <i>Judicial Role</i>	188
	3. <i>Agency Discipline</i>	188
	4. <i>The Popularity of Complex Standards of Review</i>	189
	<i>B. The Implications of Reasonableness</i>	190

INTRODUCTION

Summarizing the doctrine of judicial review in administrative law is no easy task, and in this Article, I argue that it distracts courts, scholars, and agencies from what they really do. In theory, courts must discern among and employ six different standards of review, depending on the type of agency decision challenged. Review of agency factfinding in many cases is done under a “substantial evidence” standard, but in others under an “arbitrary and capricious” standard.¹ Review of agency legal determinations triggers either so-called *Chevron* or *Skidmore* deference, or sometimes no deference at all—such as when an agency is interpreting the Constitution or the Administrative Procedure Act (APA).² Finally, courts must perform a general arbitrariness review in every case, under which, regardless of the factual conclusions or legal interpretations made by the agency, courts take a “hard look” at the agency decision to see whether the agency has sufficiently ex-

¹ See 5 U.S.C. § 706(2)(A), (E), (F) (2006) (spelling out these standards).

² For the two tests of review of law, see *Chevron, U.S.A., Inc. v. Natural Res. Def. Council, Inc.*, 467 U.S. 837, 842–43 (1984) and *Skidmore v. Swift & Co.*, 323 U.S. 134, 140 (1944). For an explanation of the operation of these tests, see *infra* Section I.A.

plained its decision and whether the decision is basically rational, based on a review of the record as a whole.³

One would assume that the differences among the various standards are consequential, and the Supreme Court has suggested exactly that: “[t]he upshot in terms of judicial review is some practical difference in outcome depending upon which standard is used.”⁴ And to observers like Paul Verkuil, these conclusions make sense. After all, “one might reasonably expect that Congress wants outcomes, defined in terms of affirmances, remands, and reversals of agency actions, to vary according to the scope of review standard chosen.”⁵

What if it turns out that these different standards of review do not make any difference in the outcome of actual cases? Although scholars and courts have not come to grips with it, administrative law outcomes do not depend on hard looks, substantial evidence, or distinctions between *Chevron* and *Mead*.⁶ Amid all the chaff of standard of review doctrine, the wheat lies in the reasonableness of the agency’s action. In fact, the “reasonable agency” standard is, increasingly clearly, the standard that courts *actually* apply to all exercises of judicial review of administrative action, no matter what standard they purport to use. Conversely, unreasonable agency action—which perhaps might be understood as agency negligence—is what courts look for when contemplating reversal.

The outcomes of actual cases prove the point: whether the question is one of fact, law, or arbitrariness, whether the agency procedures were formal or informal, whether judicial deference is required or not, the courts—even though in theory they would apply different degrees of scrutiny to each of these questions—reverse agencies slightly less than one third of the time.⁷

For the beleaguered appellate judge, who must determine the appropriate standard of review as a first step toward resolving any case, this may come as heartening news and indeed may offer a

³ *Motor Vehicle Mfrs. Ass’n v. State Farm Mut. Auto. Ins. Co.*, 463 U.S. 29, 42–43 (1983); see also Harold Leventhal, *Environmental Decisionmaking and the Role of the Courts*, 122 U. Pa. L. Rev. 509, 514 (1974).

⁴ *Dickinson v. Zurko*, 527 U.S. 150, 162 (1999).

⁵ Paul R. Verkuil, *An Outcomes Analysis of Scope of Review Standards*, 44 Wm. & Mary L. Rev. 679, 682 (2002).

⁶ The *Chevron* and *Mead* tests are discussed *infra* Section I.A.

⁷ See *infra* Section III.A.

real-world basis for a critical reevaluation of administrative procedure. Although current doctrine requires it, it has never been easy for courts to distinguish among questions of law, questions of fact, and mixed questions of law and fact; then to apply one of the various standards of review;⁸ and finally to include a catch-all review for arbitrariness. Accordingly, some introductory administrative law texts acknowledge that the judicial review standards “do not fit any neat categorization”⁹ and that there are “serious questions” about whether they “make[] any sense.”¹⁰

Lawyers, scholars, and agencies might find administrative law more predictable if they recognize that the courts do not, in the end, discern the differences among these various doctrines, frequently do not distinguish among the doctrines in application, and probably do not really care which standard of review they apply most of the time. In fact, they may wish to embrace Ernest Gellhorn and Glen Robinson’s admonition that “the rules governing judicial review have no more substance at the core than a seedless grape.”¹¹

This uncertainty is not a reason for doctrinal despair, but instead an opportunity to simplify administrative law. Administrative law would benefit from a blanket “reasonable agency” standard. Doing so would acknowledge a real-world conflation of various adminis-

⁸ The applicable standard of review is dependent on the type of procedure the agency has used. This, however, is not the only consideration. The deference owed to administrative agencies for their legal interpretations requires an inquiry into whether agency rulemaking has been “legislative” or “interpretive.” Other agency rules are exempt from review if they are “procedural” rather than “substantive,” a distinction that has bedeviled lawyers and civil procedure scholars in other areas since *Erie Railroad Co. v. Tompkins*, 304 U.S. 64 (1938).

⁹ Jerry L. Mashaw, Richard A. Merrill & Peter M. Shane, *Administrative Law: The American Public Law System* 798 (6th ed. 2009).

¹⁰ Gary Lawson, *Federal Administrative Law* 364 (4th ed. 2007).

¹¹ Ernest Gellhorn & Glen O. Robinson, *Perspectives on Administrative Law*, 75 *Colum. L. Rev.* 771, 780 (1975); see also Thomas O. McGarity, *Some Thoughts on “Deossifying” the Rulemaking Process*, 41 *Duke L.J.* 1385, 1453 (1992) (concluding that in complex scientific cases, courts apply a pass/fail standard to grade the agency); Verkuil, *supra* note 5, at 682 (“[R]eviewing judges are still struggling to make sense of these standards, especially as they apply to scope of review of facts or of law and policy.”); Patricia M. Wald, *Judicial Review in Midpassage: The Uneasy Partnership Between Courts and Agencies Plays On*, 32 *Tulsa L.J.* 221, 258 (1996) (“After fifty years . . . we have yet to agree on how this review should operate in practice. We are still struggling with where to draw the line between obsequious deference and intrusive scrutiny.”).

trative law standards of review—and not just by result, but also by increasingly creaky and oft-analogized doctrine that appears to be headed toward reasonableness anyway.

The standard would also put the point of entry for the relationship between courts and agencies on firmer ground. It would simplify the requirements for judicial review and make the law more accessible to the laity.

Even more broadly, reasonableness review would illustrate how, in administrative law, appellate judges play the role typically assumed by jurors, trial judges, and magistrates, all of whom are asked to find whether challenged conduct—ranging from negligence actions in tort to searches and seizures in criminal law—is reasonable.¹² Appellate judges, whatever their expertise in legal analysis, need not be thought of so differently. As with these other cases, appellate review of agency action is a point of entry into the legal system for a wide array of facts, and it makes sense that reasonableness should underpin these broad gateways to adjudication.

Finally, a reasonableness standard might allow lawyers and judges to concentrate on the real-world issues that probably affect judicial decisionmaking more than do the standards of review. Administrative law may often (but not always) be more about the politics of the dispute, the specifics of agency conduct, and—especially in the D.C. Circuit—the application of community standards and legal principles to the policies of the administrative state, than about standard of review.¹³ These factors, of course, are the

¹² In fact, the reasonableness standard that courts actually apply in administrative law may apply even more broadly. See, e.g., Martin Shapiro, *The Giving Reasons Requirement*, 1992 U. Chi. Legal F. 179, 189 (“[F]rom a reviewing court’s perspective, in American constitutional law and all other constitutional law, the ultimate test of a statute or other government action is reasonableness.”). For the classic statement of the role of reasonableness in Fourth Amendment jurisprudence, see *Terry v. Ohio*, 392 U.S. 1, 21 (1968) (“The scheme of the Fourth Amendment becomes meaningful only when it is assured that at some point the conduct of those charged with enforcing the laws can be subjected to the more detached, neutral scrutiny of a judge who must evaluate the reasonableness of a particular search or seizure in light of the particular circumstances.”).

¹³ See Peter M. Haas, *Epistemic Communities and International Policy Coordination*, Introduction to *Knowledge, Power, and International Policy Coordination*, 46 Int’l Org. (Special Issue) 1, 3 (Peter M. Haas ed., 1992) for a well-known review of the importance of these elite communities in policymaking.

same as those thought to influence juries in negligence cases.¹⁴ Making the standard of review a complicated opening question needlessly obscures matters of real-world import, for no good purpose.

In this Article I begin with an appraisal of the doctrine—or rather doctrines—of judicial review and then describe how, if taken seriously as doctrine and matters of logic, their practical resemblance is drawing ever closer. Part II of the Article draws on doctrinal scholarship that has addressed narrow questions of administrative law, such as whether one particular administrative law standard resembles another. Scholars and courts have, as matters of logic and doctrinal practice, concluded that in many cases this administrative law standard does indeed look a lot like that one, and in some cases, the differences among various standards of review are not at all clear. Moreover, I identify a developing convergence of the standards of judicial review on the second step of the famed *Chevron* two-step standard of review.¹⁵ That standard inquires whether an agency's interpretation of its legal authority was reasonable.¹⁶ From a series of discrete comparisons, I infer a developing consensus that the various standards, in the end, look to reasonableness as the baseline for evaluating agency action.

This logical account is bolstered by quantitative analyses of judicial review of administrative action, which I address in Part III of this Article. Prior empirical studies of such review have suggested that, uncannily, across a variety of standards, across a variety of discrete issue areas, post-*Chevron* courts reverse agencies approximately one third of the time.¹⁷

These studies have never been aggregated to draw broader conclusions about whether varying standards of judicial review actu-

¹⁴ And have been thought to do so since Oliver Wendell Holmes' common law lectures. See O.W. Holmes, Jr., *The Common Law* 111 (1881) (suggesting that juries are best suited to determine liability under negligence standards because of their familiarity with community standards); see also Catharine Pierce Wells, *Tort Law as Corrective Justice: A Pragmatic Justification for Jury Adjudication*, 88 *Mich. L. Rev.* 2348, 2411 (1990) (arguing that jury adjudication is important because it requires jurors to strive to apply community standards rather than subjective judgment in negligence cases).

¹⁵ *Chevron, U.S.A., Inc. v. Natural Res. Def. Council, Inc.*, 467 U.S. 837, 842–43 (1984).

¹⁶ *Id.* at 844.

¹⁷ See *infra* Section III.A.

ally affect whether agencies win or lose cases. Moreover, they lack an empirical review of one final variant: no one has yet looked comprehensively at judicial review of agencies' formal findings of fact. The substantial-evidence standard is one of the principal ways that courts evaluate agency action and is always taught as an essential part of judicial review in introductory administrative law classes.¹⁸ This Article remedies that gap in the literature by adding that final quantitative analysis. It examines agency validation rates in substantial-evidence review cases in the D.C. Circuit between 2000 and 2004. The dates are chosen because they are roughly contemporaneous with a number of other empirical studies of other administrative law standards of review.

I find that in substantial-evidence review, too, appellate courts reverse agency formal adjudications of fact slightly more than one third of the time. Moreover, although prior scholarship has focused on the ideological components of judging to explain who wins, my analysis of those components in substantial evidence—at least for the data collected here—does not suggest a statistically significant relationship between the party of the judge, the nature of the agency, or a number of other variables, and the agency's validation rate. A single study does not cast much doubt on the importance of ideology in adjudication—which has been found in a number of other studies of judicial review of agency action, and which I and most legal scholars have long believed to be plausible.¹⁹ It does, however, provide some support for the larger method adopted here, which is to aggregate validation rates—rather than to scrutinize selected cases—and tease out a model with legal and ideological variables to see how they work.

The empirical strategies chosen are a weighted average of my study and prior studies—since aggregated studies, like political

¹⁸ I am unaware, at least, of any administrative law casebook that dispenses with it. Moreover, the case announcing the test for substantial evidence review, *Universal Camera Corporation v. NLRB*, 340 U.S. 474 (1951), was, as of 2006, the second most cited case in administrative law. See David Zaring, *The Most Cited Cases in Administrative Law*, *Concurring Opinions*, Mar. 29, 2006, http://www.concurringopinions.com/archives/2006/03/the_most_cited.html.

¹⁹ See *infra* Section III.B for a review of these studies.

polls, are more reliable than a single study²⁰—and, for the substantial evidence study, descriptive statistics and simple logistic regression.

Caveats are important in evaluating these quantitative analyses: selection bias is a persistent problem for studies of judicial review because litigants decide which cases to bring, and this study would be no better than others if it tried to tease out insights about the power of agencies or the supine *vel non* nature of judicial review of the administrative state.²¹ But that is not the goal here; the comparison is not between courts and agencies, nor between liberal and conservative judges; it is instead among various standards of review. It is still true that different cases may be brought under different standards of review; there may be a set of selection biases at work here, so caution is warranted. Nonetheless, if, regardless of the different standards of review, courts reverse agencies at the same rate, all the doctrinal complexities and curlicues in this area of administrative law bear reconsideration.

This Article concludes with that reconsideration. I consider the implications of a reasonable-agency standard, and identify some additional doctrinal confusions that could be eliminated by its adoption. I explore what the emerging prominence of the reasonableness standard—and the recommended explicit adoption of it—would mean for the judicial role of administrative law, which has been based on an expertise model that is not obviously consistent with reasonableness. In the end, it could be that the emerging adoption of reasonableness in practice means that efforts to draw fine distinctions about judicial review—much a province of traditional administrative law—are doomed to futility, suggesting yet again that Oliver Wendell Holmes was right when he claimed that “[t]he life of the law has not been logic: it has been experience.”²²

²⁰ For an example of a pooled weighted average of studies, see Lucian Bebchuk, Alma Cohen & Allen Ferrell, Does the Evidence Favor State Competition in Corporate Law?, 90 Cal. L. Rev. 1775, 1791–94 (2002).

²¹ For a discussion of selection bias in administrative law, see Cary Coglianese, Empirical Analysis and Administrative Law, 2002 U. Ill. L. Rev. 1111, 1118–19.

²² Holmes, *supra* note 14, at 1.

I. STANDARD OF REVIEW DOCTRINE

Although courts tend to reverse agencies at the same rate regardless of the standard of review they apply, those standards are doctrinally distinct. Functionally, they distinguish between whether agencies were finding facts or interpreting law, the procedures used by the agency, and, if a legal interpretation is involved, the source of law upon which the agency is basing its action. Finally, courts apply a catch-all arbitrariness standard to all agency action. The result is a rather bewildering array of standards of review, standards that the puissant administrative lawyer must be able to discern and cite in her challenge to, or defense of, agency action.²³ Because this Article's point is that these various doctrinal standards are not very consequential, either as a matter of theory or as a matter of practice, those familiar with these standards may wish to skip this section. But because courts usually (although, as we will see, not always) announce that a particular standard governs their review, it is worth taking them at their word—presuming that the doctrinal distinctions are coherent—and explaining what it is the various standards are supposed to do as a matter of legal analysis.

A. Review of Law

Agency interpretations of law are, if reviewable at all,²⁴ reviewable under one of three basic standards: the deferential *Chevron*

²³ It is also worth noting that, to make things more confusing, none of these standards is duplicated by the standards of review used by appellate courts to review trial courts. Of course—as might be expected given the analysis here—it is by no means clear that these doctrinal differences are strictly observed. The Supreme Court tried to compare the agency review standard of substantial evidence with the judicial review standard of clear and convincing evidence and concluded that

[t]he court/agency standard, as we have said, is somewhat less strict than the court/court standard. But the difference is a subtle one—so fine that (apart from the present case) we have failed to uncover a single instance in which a reviewing court conceded that use of one standard rather than the other would in fact have produced a different outcome.

Dickinson v. Zurko, 527 U.S. 150, 162–63 (1999); see also Lawson, *supra* note 10, at 363 (“For the most part, the contemporary system of judicial review of federal administrative agency action employs *none* of the benchmarks familiar from appellate review of lower court decisionmaking.”).

²⁴ Some agency legal determinations, such as those related to national security, are exempt from review under the APA, 5 U.S.C. § 552 (2006), while other decisions are “committed to agency discretion by law.” *Id.* § 701.

standard, the less deferential *Skidmore* standard—which applies when the test set forth in *United States v. Mead* is met—and the no-deference-at-all standard of de novo review.²⁵ These tests apply depending on the source of law that the agency uses to justify the action being disputed in court.

Chevron announces the standard most often applied by reviewing courts; indeed, it is the most cited case in administrative law.²⁶ The standard applies when an agency is interpreting a statute it has been charged with administering through some order imbued with the force of law. The standard of review under *Chevron* consists of two steps. For the first step, the reviewing court must ask whether, after “employing traditional tools of statutory construction,” it is evident that “Congress has directly spoken to the precise question at issue.”²⁷ If so, the statute is “unambiguous[],” and the agency must not differ from Congress’ clearly expressed command.²⁸ If, however, the court decides that the statute is ambiguous, it then moves to step two of the inquiry. That step requires the court to uphold the agency’s interpretation so long as it is “based on a permissible construction of the statute.”²⁹ Appellate courts have interpreted this to require deference to any reasonable interpretation of the statute offered by the agency.

²⁵ A final basis for agency action involves an agency’s interpretation of its own regulations, to which reviewing courts confer *Auer*, or *Seminole Rock*, deference. The *Auer* test provides that if an agency applies and interprets a test that is “a creature of [its] own regulations, [its] interpretation of it is, under our jurisprudence, controlling unless ‘plainly erroneous or inconsistent with the regulation.’” *Auer v. Robbins*, 519 U.S. 452, 461 (1997) (quoting *Bowles v. Seminole Rock & Sand Co.*, 325 U.S. 410, 414 (1945)). Although the point is debatable, many scholars believe that *Auer/Seminole Rock* deference is indistinguishable from *Chevron* deference. “It would seem that there are few, if any, cases in which the standard applicable under *Chevron* would yield a different result than the ‘plainly erroneous or inconsistent’ standard set forth in *Bowles v. Seminole Rock & Sand Co.*” *Paralyzed Veterans of Am. v. D.C. Arena L.P.*, 117 F.3d 579, 584 (D.C. Cir. 1997). Partly for that reason, this Article will not consider *Auer* deference separately—though the interpretive task in *Auer* is a bit different: rather than interpreting some source of law that comes from outside the agency, regulatory interpretation requires it to interpret law that the agency itself has made.

²⁶ At least, as of 2006. See Zaring, *supra* note 18.

²⁷ *Chevron, U.S.A., Inc. v. Natural Res. Def. Council, Inc.*, 467 U.S. 837, 842, 843 n.9 (1984).

²⁸ *Id.* at 842–43.

²⁹ *Id.* at 843.

Chevron is highly cited partly because when agencies act, they usually do so on the basis of a statutory mandate given to them by Congress, authority that they identify in the action under review. Of course, sometimes mandates to agencies may be very broad, as when Congress instructs the agency to protect the “public interest” in broadcast communications,³⁰ for example, or to ensure that any “occupational safety and health standard” promulgated by the Occupational Safety and Health Administration (OSHA) is “reasonably necessary or appropriate to provide safe or healthful employment and places of employment.”³¹ Interpreting the limits of very broad mandates can be difficult, even to the point of raising separation of powers and other constitutional issues.³² In other cases, the mandates may be very specific, as when Congress directs an agency to, say, promulgate new regulations by a particular date.³³ In such cases, Congressional intent is quite clear, and the *Chevron* exercise is presumably an easy one.

Quite confusingly, however, *Chevron* is not the standard that applies in every case where an agency is interpreting a statute it administers, because sometimes the agency does so in a case where it is not acting with “force of law,” as it might do if it were preparing materials for a handbook designed to educate its employees about its mission. In those and other cases where the agency is not acting with force of law, *Chevron* deference does not apply. The Supreme Court has explained that “administrative implementation of a particular statutory provision qualifies for *Chevron* deference 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.”³⁴ This test, set forth in *United States v.*

³⁰ Communications Act of 1934, 47 U.S.C. § 201(b) (2006) (empowering the Commission to prescribe rules and regulations “in the public interest”).

³¹ 29 U.S.C. § 652(8) (2006).

³² For example, Cass Sunstein believes that the extraordinarily broad OSHA mandate implicates the nondelegation doctrine. Cass R. Sunstein, *Is OSHA Unconstitutional?*, 94 Va. L. Rev. 1407, 1407–08 (2008).

³³ The Sarbanes-Oxley Act and the Foreign Investment and National Security Act (FINSA) are examples of such statutes. See Sarbanes-Oxley Act, Pub. L. No. 107-204, § 208, 116 Stat. 745, 775 (2002); *id.* § 307 at 784; *id.* § 401(b) at 786; Foreign Investment and National Security Act of 2007, Pub. L. No. 110-49, § 9, 121 Stat. 246, 259 (2007).

³⁴ *United States v. Mead Corp.*, 533 U.S. 218, 226–27 (2001).

Mead, requires reviewing courts to, effectively, engage in a “*Chevron* [s]tep [z]ero” inquiry into whether the *Chevron* test should be applied at all.³⁵

The *Mead* decision provides that agencies may show that they were empowered to act with the force of law, and are accordingly entitled to receive *Chevron* deference, “in a variety of ways, as by an agency’s power to engage in adjudication or notice-and-comment rulemaking, or by some other indication of a comparable congressional intent.”³⁶

Where an agency is unable to show that it has acted in these ways, it receives so-called *Skidmore* deference, which means that the agency’s statutory interpretations are “not controlling upon the courts.”³⁷ *Skidmore v. Swift* concluded that the weight accorded an agency interpretation “in a particular case will depend upon the thoroughness evident in its consideration, the validity of its reasoning, its consistency with earlier and later pronouncements, and all those factors which give it power to persuade.”³⁸ The putative difference is that, in a *Chevron* case, a court *must* defer to a reasonable interpretation of a statute by an agency, while in a *Mead/Skidmore* context it *may* defer based on how persuasive it finds the agency interpretation. Distinguishing between *Chevron* and *Mead* contexts has become a preoccupation of appellate courts; between the announcement of the test on June 18, 2001, and July 31, 2009, federal courts of appeals cited to *Mead* 440 times.

De novo review is appropriate when agencies are interpreting laws that they do not have a special responsibility to administer, like the Constitution, the APA, or Title VII.³⁹ It also applies in some other rare contexts, such as when the agency action is adjudi-

³⁵ Cass R. Sunstein, *Chevron Step Zero*, 92 Va. L. Rev. 187, 191 (2006).

³⁶ *Mead*, 533 U.S. at 227.

³⁷ *Skidmore v. Swift & Co.*, 323 U.S. 134, 140 (1944). For an example of such a statute, see Title VII of the 1964 Civil Rights Act, 42 U.S.C. § 2000e-5 et seq. (2000).

³⁸ *Skidmore*, 323 U.S. at 140.

³⁹ *Crowell v. Benson*, which first approved the adjudication of private rights by an administrative agency rather than an Article III court, is an example of a case where facts necessary to adjudication of a constitutional right were subject to de novo review. 285 U.S. 22, 46 (1932). The APA and Title VII apply to *all* federal agencies (even the Equal Employment Opportunity Commission), thus leaving none of them with the power to interpret the test authoritatively. See 5 U.S.C. §§ 551–552 (2006); 42 U.S.C. § 2000e(a) (2006).

catory in nature and the agency factfinding procedures are inadequate, or when issues that were not before the agency are raised in a proceeding to enforce nonadjudicatory agency action.⁴⁰ It is not a particularly common standard of review, but it—in theory—represents the few cases where particularly suspicious procedures result in agency action or where agency action invokes universal legal principles as the basis of its action.

B. Review of Fact

Agency determinations of fact are subject to a different set of tests during review by the courts; these determinations are made when courts review agency conclusions about the world, rather than about the agency's legal authority to act in the world. The nature of this review depends upon whether the agency found facts through formal or informal proceedings, in keeping with a distinction made by observers almost immediately after the passage of the APA, although only hinted at in the APA itself.⁴¹ Fact-based review thus adds two more standards of review to the three that apply for agency legal determinations.

Agency factfinding through formal proceedings are those made “on the record.”⁴² Indeed, agencies tend to adopt formal proceedings only when the words “on the record” appear in a statute setting forth their adjudicatory obligations.⁴³ These proceedings look

⁴⁰ 5 U.S.C. § 706(2)(F) (2006).

⁴¹ The APA specifies elaborate procedures for formal adjudications, but neither specifies processes for informal adjudications nor uses the term. See 5 U.S.C. §§ 554, 556–557 (containing the APA's adjudication requirements); *id.* § 555(e) (requiring that “[p]rompt notice shall be given of the denial in whole or in part of a written application, petition, or other request of an interested person made in connection with any agency proceeding. Except in affirming a prior denial or when the denial is self-explanatory, the notice shall be accompanied by a brief statement of the grounds for denial”). Although this Article is concerned with modern standards of judicial review, it is worth noting that in the pre-APA era “mixed” questions of law and fact are entitled to deference similar to “pure” fact questions under the appropriate scope of review standards. See, e.g., *NLRB v. Hearst Publ'ns*, 322 U.S. 111, 130–31 (1944).

⁴² 5 U.S.C. §§ 556(c), 557(b)(2).

⁴³ Though the ability of agencies to choose the level of formality of their adjudication procedures has been debated by the courts for some time. Compare *Dominion Energy Brayton Point, LLC v. Johnson*, 443 F.3d 12, 18–19 (1st Cir. 2006) (finding that agencies may choose whether to conduct formal or informal adjudications if their governing statute does not unambiguously require formal proceedings), and *Chem. Waste Mgmt. v. U.S. Env'tl. Prot. Agency*, 873 F.2d 1477, 1482 (D.C. Cir. 1989)

like trials and are governed by Sections 556 and 557 of the APA.⁴⁴ They require an administrative law judge (ALJ) to preside over the taking of evidence, permit the parties to the proceeding to present their cases, cross-examine opposing witnesses, and require the preparation of a transcript, among other things.⁴⁵ When the agency finds facts through these procedures, Section 706(2)(E) of the APA provides that courts may reverse their findings only if “un-supported by substantial evidence.”⁴⁶ In *Universal Camera v. NLRB*, the Court interpreted substantial evidence review as an inquiry into whether the facts found were supported by “such relevant evidence as a reasonable mind might accept as adequate to support a conclusion.”⁴⁷ Put another way, “substantial evidence” exists if “a reasonable jury could have found” the facts the agency found.⁴⁸

Formal procedures may occur whenever an ALJ is required by a statute to conduct a hearing before the agency may take action. For example, formal procedures apply when the Social Security Administration attempts to remove a disability recipient from the rolls, because the governing statute requires formal adjudication as to whether the disability benefits recipient continues to be disabled.⁴⁹ Labor cases that turn on factfinding also ordinarily are conducted through formal adjudication, also because the statute requires as much.⁵⁰ For example, *Universal Camera*—which provided the Supreme Court with the opportunity to articulate what review of formal adjudication is required—concerned judicial review of the NLRB’s weighing of evidence in determining whether a

(same), with *Seacoast Anti-Pollution League v. Costle*, 572 F.2d 872, 877 (1st Cir. 1978) (finding that if the statute did not specify the proceedings required, formal proceedings would be presumed).

⁴⁴ 5 U.S.C. §§ 556–57.

⁴⁵ *Id.* § 556.

⁴⁶ 5 U.S.C. § 706(2)(E) (2006).

⁴⁷ 340 U.S. 474, 477 (1951) (quoting *Consol. Edison Co. v. Labor Bd.*, 305 U.S. 197, 229 (1938)).

⁴⁸ *Allentown Mack Sales & Serv. v. NLRB*, 522 U.S. 359, 366–67 (1998).

⁴⁹ For an overview, see Office of Disability & Income Sec. Programs, Soc. Sec. Admin., *The Red Book, A Summary Guide to Employment Supports for Individuals with Disabilities Under the Social Security Disability Insurance and Supplemental Security Income Programs*, SSA Pub. No. 64-030 (2008).

⁵⁰ 29 U.S.C. § 160(f) (2006).

worker was fired because he supported unionization or because of his job performance.⁵¹

Informal APA proceedings are not required to be on the record. They cover a panoply of administrative action where facts are found without the involvement of an ALJ. Agencies conduct informal adjudications whenever the statute does not contain the words “on the record” in elaborating their adjudicatory responsibilities. Informal adjudications range from determinations by the National Park Service as to whether someone should get a camping permit (considering factors such as whether a particular campsite is oversubscribed, whether the applicant paid the entrance fee, and so on) to rather more technical analysis, such as the effect on water ecosystems of a discharge permit granted to a nuclear power plant.⁵² Facts found during the course of notice-and-comment rulemaking also receive this standard of review, as might be the case when an agency is determining what levels of particulate matter in the atmosphere pose a threat to human health.⁵³

A court may overturn informal factfinding on review if it was done in an “arbitrary” or “capricious” manner, a standard that also applies to arbitrariness review.⁵⁴ In the world of factfinding, to satisfy that standard of review, the “inquiry into the facts is to be searching and careful,” but “the ultimate standard of review is a narrow one.”⁵⁵ It is a “substantial inquiry” and “a thorough, prob-

⁵¹ *Universal Camera*, 340 U.S. at 476–78.

⁵² The latter question was presented by two well-known First Circuit cases on formal adjudication. See *Seacoast Anti-Pollution League v. Costle*, 572 F.2d 872, 874 (1st Cir. 1978), abrogated by *Dominion Energy Brayton Point, LLC v. Johnson*, 443 F.3d 12 (1st Cir. 2006).

⁵³ Particulate matter disputes, because they offer the prospect of real expense to much of the American economy, are particularly likely to be contested. The Clinton administration’s dispute eventually reached the Supreme Court in *Whitman v. American Trucking Associations*, 531 U.S. 457, 480–81 (2001); the Bush administration’s efforts were reversed by the D.C. Circuit in *American Farm Bureau Federation v. EPA*, 559 F.3d 512, 515 (D.C. Cir. 2009).

⁵⁴ 5 U.S.C. § 706(2)(A) (2006). For a leading discussion of the requirements of judicial review of informal factfinding, see *Association of Data Processing Service Organizations v. Board of Governors of the Federal Reserve*, 745 F.2d 677, 681–86 (D.C. Cir. 1984) (Scalia, J.).

⁵⁵ *Citizens to Pres. Overton Park, Inc. v. Volpe*, 401 U.S. 402, 416 (1971); see also Paul R. Verkuil, *Judicial Review of Informal Rulemaking: Waiting for Vermont Yankee II*, 55 Tul. L. Rev. 418, 424 (1981) (characterizing the *Overton Park* approach as “intrusive substantive review”).

ing, in-depth review.”⁵⁶ This classic formulation—set forth in *Citizens to Preserve Overton Park v. Volpe*—includes language that is both strict and relaxed, suggesting some ambiguity about what precisely the arbitrary and capricious standard for review of fact requires. Nevertheless, courts have repeatedly turned to the case for doctrinal guidance about what, exactly, they are supposed to do when evaluating facts found by agencies in informal proceedings.

In all of these matters, courts have repeatedly complained that understanding what exactly the applicable standard of review of agency factfinding requires of them is by no means clear. All of this consolidation comes against a background of confusion about what, exactly, the two standards of review actually mean. Judges have complained that “the substantial evidence standard sounds simple enough. But for appellate courts charged with reviewing the decisions of administrative tribunals, applying that standard has proved to be a veritable will-o’-the-wisp.”⁵⁷ Similar protests have been made for arbitrary and capricious review of agency factfinding.⁵⁸

C. Arbitrariness Review

There is a final standard of judicial review that applies to all agency action and, confusingly, is based on the same part of the APA that is used to provide one of the factfinding standards of review, the “arbitrary and capricious” language of the statute.⁵⁹ The Supreme Court has interpreted this provision to require that every agency is subject to a baseline standard of review of arbitrariness, the terms of which were set out in the Court’s decision in *Motor Vehicle Manufacturers Association v. State Farm Insurance Co.*⁶⁰ *State Farm* review turns on whether, in doing whatever the agency did—be it finding facts, interpreting statutes, or resolving mixed questions of fact or law—the agency has considered the relevant factors or has based its decision on factors it is not supposed to

⁵⁶ *Overton Park*, 401 U.S. at 415.

⁵⁷ *Tex. World Serv. Co. v. NLRB*, 928 F.2d 1426, 1430 (5th Cir. 1991).

⁵⁸ See, e.g., Steven J. Cann, *Administrative Law* 153 (4th ed. 2006) (“Unfortunately, describing what the arbitrary and capricious standard applies to is easier than describing the standard and how it is applied.”).

⁵⁹ 5 U.S.C. § 706(2)(a).

⁶⁰ 463 U.S. 29, 42–44 (1983).

consider and “whether there has been a clear error of judgment.”⁶¹ As the Court held in *State Farm*, the agency “must examine the relevant data and articulate a satisfactory explanation for its action including a ‘rational connection between the facts found and the choice made’” to meet this test.⁶²

Arbitrariness review applies in all cases, and it has variously been thought of as a constraint on the substantive rationality of the agency’s decisionmaking or, alternatively, as something a bit more procedural, requiring the agency to give reasons for the decision it made. Examples of reversals for substantive irrationality are easy to hypothesize, though less easy to find in the real world of agency action. Suppose, however, that the Environmental Protection Agency (EPA) were to ban automobiles because of their effects on the environment. Such an order might lie within EPA’s statutory authority and even might be justified by a long record of facts found about the environmental consequences of roadbuilding, the combustion engine, and so on. But the order might nonetheless be too extreme for a court to countenance it; if so, it would presumably be reversed on the basis of application of the *State Farm* arbitrariness review.

Far easier to come by are examples of cases where the agency failed to meet its reason-giving requirement—that is, to explain how the legal authority and conclusions about the world manifest themselves in the particular policy announced by the agency. Still, one often suspects that judicial faulting of agency reason-giving may be difficult to distinguish from judicial disapproval of the substantive rationality of the agency decision. In the *State Farm* case itself, the Supreme Court reversed the National Highway Transportation Safety Administration for repealing its own early-1980s decision requiring cars to be equipped with air bags and seat belts without explaining why it changed its view that these safety features were necessary.⁶³

⁶¹ Id. at 43 (citing *Bowman Transp., Inc. v. Ark.-Best Freight Sys.*, 419 U.S. 281, 285 (1974)).

⁶² Id. (quoting *Burlington Truck Lines v. United States*, 371 U.S. 156, 168 (1962)).

⁶³ Kathryn Watts believes that the problem with these decisions is that the explanation for policy changes—a change in political administration, frequently—are not acknowledged to be legitimate ones in arbitrariness review. Kathryn A. Watts, *Proposing a Place for Politics in Arbitrary-and-Capricious Review*, 119 *Yale L.J.* 2, 7–8 (2009).

D. Implications of the Six Standards of Review

The variety of the tests I have described above, and the difficulties of deciding which apply when, has often made sorting through the standards of review an exceedingly complicated task. Appellate courts have protested that distinguishing among the standards is “fraught with ambiguity.”⁶⁴

It is easy to sympathize with the perplexed reaction of the courts, especially where degrees of deference are concerned. Yet discerning which degree of deference applies to the various legal interpretations by agencies is only part of the problem. Distinguishing between questions of law and questions of fact has never been easy either, at least at the margin. Furthermore, the very prospect of unpacking the difference between “unreasonableness” in review of agency interpretations of law, as *Chevron* step two requires, and “arbitrariness” in review of factual determinations and policy choices, as *State Farm* requires, suggests the complexities of the inquiry in its own right.

When faced with such doctrinal challenges—a particularly complicated structure of review combining no fewer than six different tests based on the type of agency action being reviewed and the task of determining the precise thing that the agency did, whether it involved factfinding, legal interpretation, or policymaking—it is worth taking a step back and asking whether the tests obscure more than they reveal.

Does the elaborate scheme of judicial review of agency action matter? The administrative law casebooks are full of sympathetic bows to the student learner, acknowledging that there are all these standards of review out there and that they seem to be difficult to

⁶⁴The Ninth Circuit has said as much when it comes to distinguishing between *Chevron* and *Mead*. See *United States v. W.R. Grace & Co.*, 429 F.3d 1224, 1235 (9th Cir. 2005) (“Following *Mead*, the continuum of agency deference has been fraught with ambiguity. Our decisions understandably have been conflicted as to whether *Chevron* deference only applies upon formal rulemaking and whether lesser deference applies in other situations.”) (citations omitted); see also *Wilderness Soc’y v. U.S. Fish & Wildlife Serv.*, 316 F.3d 913, 921 (9th Cir. 2003) (“After *Mead*, we are certain of only two things about the continuum of deference owed to agency decisions: *Chevron* provides an example of when *Chevron* deference applies, and *Mead* provides an example of when it does not.”).

distinguish.⁶⁵ Unfortunately, however, as these casebooks suggest, the sundry tests must ultimately be learned in order to grasp a critical issue in administrative law: what courts do to agencies. And if law students are doomed to struggle with the complexities of administrative law, one can only imagine how it would be for the laypeople.

II. THE REALITY OF ADMINISTRATIVE REVIEW

Can administrative law really be this complicated? A variety of administrative law scholars and courts have had their doubts. These observers have analogized one standard of review to another, and in the end concluded that they are pretty similar. The D.C. Circuit once concluded that, as among the three standards of review of legal questions, “the result is the same whether the court applies *de novo* review, deference under *Skidmore v. Swift & Co.*, or *Chevron* deference.”⁶⁶ With regard to agency adjudications, as William Funk and Richard Seamon have observed, “[g]iven the deference . . . accorded an agency’s interpretation made in an adjudicatory application of a law and the deference accorded an agency’s determination of a mixed question of law and fact under substantial evidence review, the *outcomes are likely to be the same no matter which review standard is used.*”⁶⁷

⁶⁵ See, e.g., Ernest Gellhorn & Clark Byse, *Administrative Law* 1037 (Peter Strauss et al. eds., 10th ed. 2003) (“Can you put *Chevron* together with *State Farm* and *Overton Park*?”); *id.* at 1034; *Administrative Law And Regulatory Policy* 415 (Stephen Breyer et al. eds., 5th ed. 2002) (noting the “many puzzles” involved with sorting through the standards); see also Ernest Gellhorn, Justice Breyer on Statutory Review and Interpretation, 8 *Admin. L.J. Am. U.* 755, 755 n.4 (questioning “whether the legal rules were worth serious study—or at least the amount of time usually invested in them in the classroom or casebooks”).

⁶⁶ *Bullcreek v. Nuclear Regulatory Comm’n*, 359 F.3d 536, 541 (D.C. Cir. 2004) (citation omitted).

⁶⁷ William F. Funk & Richard H. Seamon, *Examples & Explanations: Administrative Law* 310 (3rd ed. 2009) (emphasis added). The task is made all the more difficult by the fact that none of these standards of review are, at least in theory, matched by the standards of review applied to appellate review of trial courts. Lawson, *supra* note 10, at 363. Instead of *de novo* review of legal findings, agencies receive deference for the legal interpretations they are charged by Congress with making. Instead of clearly erroneous review of factfinding, agencies get substantial evidence or arbitrary-and-capricious review. And there is no judicial review analog for the hard look taken at the policy choices made by agencies.

But these sorts of comparisons have been made among specific standards of review rather than universally. I argue that, considered as a gestalt of administrative law, there is an emerging consensus of scholars and courts who reject this sort of doctrinal formalism. I build upon the logical advances of these authorities, who have suggested that one particular standard of judicial review works logically like another one. Indeed, there is a lot of “this one looks like that one” in administrative law.

One might also ask if it is really consistent with our intuitions about judging—let alone our views on how humans balance the inclination to simplify with the inclination to elaborate (and accordingly complexify)—to assume that courts bother to keep the standards of review of administrative agencies distinct. In my view, it is more likely that the reasonableness heuristic—used in so many other legal contexts, such as tort law and Fourth Amendment jurisprudence on searches and seizures—*is* hard to banish from the minds of judges. Moreover, this might be the case whether judges rarely see confusing administrative schemes (as might be the case with judges who sit on appellate courts outside of the District of Columbia) or whether they have grown so familiar with certain agencies so as to have moved beyond doctrine to look-and-feel evaluations of administrative action (as might be the case for judges on the D.C. Circuit).

This Part is a synthetic one; it draws on the insights of scholars and courts who have decided that they are unable to distinguish among the six standards of review in administrative law. If one pools these insights, the doctrine of judicial review of agency action itself seems to include a remarkable number of analogies from one standard of review to another. I argue that it makes sense to take a more holistic approach to judicial review of agency action, partly because it comports with what we suspect judges actually do and partly for normative reasons—because it would make the law more transparent and more predictable if we formally recognized the “reasonable agency” standard that courts actually apply to review of agency action.

It is worth remembering, however, that the chief goal of this Article is not to make a logical argument that, as a matter of doctrine, reasonableness trumps all. After all, one might disagree that a particular doctrine looks like another one in certain cases. For exam-

ple, de novo review has been interpreted to be quite different from deferential standards of review even by legal realists such as Richard Posner.⁶⁸ But even if one disagrees with a particular analogy used by a particular judge or law professor, one still could conclude that, in light of the striking similarities of analysis across doctrines, there is a much more uniform standard of judicial review than the existence of the six different doctrinal tests might lead one to believe.

This is not to suggest that the emerging logic of judicial review *requires* courts to treat all administrative law cases the same way—such an argument would be Langdellian in its formality, and this is not a formalist article. The point here is to challenge the idea that the six standards of judicial review have, as a matter of practice, not been kept quite as distinct as a pure doctrinalist might expect.

At any rate, even those who believe that the standards of review rest on functionally distinguishable bases will have to grapple with the remarkable similarities in reversal rates among them. If courts reverse agencies approximately one third of the time, regardless of the standard of review applied, all of the doctrinal complexities will have met a considerable real-world challenge. In the next Section of this Article, I will bolster my argument with an analysis of the quantitative research conducted in this field. To that research, I will add my own quantitative analysis of how the D.C. Circuit has conducted substantial evidence review. In this Section, I note the many occasions in which courts and scholars have found it to be particularly difficult to distinguish among various standards of review.

A. Review of Law Is Becoming Review for Reasonableness

The conflation of the various standards of review into a reasonableness inquiry is particularly clear with the most used of those standards: the *Chevron* test. Although *Chevron* has been dubbed by courts as announcing a “famous” two-step inquiry, the utility of the first step is not obvious.⁶⁹ The first step requires the court to evaluate agency legal interpretations to see if Congress clearly

⁶⁸ *Infra*, notes 169–69.

⁶⁹ Sunstein, *supra* note 35, at 190 (“*Chevron* famously creates a two-step inquiry for courts to follow in reviewing agency interpretations of law.”).

spoke to the legal question at issue. But does the first step add anything to the second step, which inquires whether the agency action was reasonable?

It is not clear why it would ever be reasonable to interpret a statute contrary to Congress' clearly expressed intent. Therefore, any agency interpretation that fails the first step of *Chevron*, which looks to congressional intent, would surely fail the second step, which turns on the reasonableness of the inquiry. In other words, there are no cases that would fail step one and not also fail step two. Conversely, fidelity to congressional intent seems like it would go a long way toward demonstrating the reasonableness of the agency's legal interpretation. So the step two inquiry surely encompasses what courts are supposed to do in step one.

For this reason, Matthew Stephenson and Adrian Vermeule believe that *Chevron* itself is better understood as a one-step test. As they say, "[t]he single question is whether the agency's construction is permissible as a matter of statutory interpretation; the two *Chevron* steps both ask this question, just in different ways. As a result, the two steps are mutually convertible."⁷⁰ And that test, in turn, is based on reasonableness. "*Chevron* calls for a single inquiry into the reasonableness of the agency's statutory interpretation," Stephenson and Vermeule argue.⁷¹ And surely it is true that administrative departures from Congress' clearly expressed intent would always be unreasonable exercises in statutory interpretation.

Although courts routinely apply the *Chevron* doctrine, actual judicial practice reveals their recognition that the two-step inquiry tends to collapse into one. The D.C. Circuit has already acknowledged that "the factors involved in the first 'step' are also pertinent to whether an agency's interpretation is 'reasonable.'"⁷² In the *Chevron* opinion itself, the Supreme Court expresses some ambivalence about its announced test, for, as the D.C. Circuit has observed, "the *Chevron* Court itself never specified which step it was applying at any point in its analysis."⁷³

⁷⁰ Matthew C. Stephenson & Adrian Vermeule, *Chevron Has Only One Step*, 95 Va. L. Rev. 597, 599 (2009).

⁷¹ *Id.* at 598.

⁷² *Sweet Home Chapter of Cmty. for a Great Or. v. Babbitt*, 30 F.3d 190, 193 (D.C. Cir. 1994), rev'd, 515 U.S. 687 (1995).

⁷³ *Id.*

I think Stephenson and Vermeule are probably correct, though it is worth noting that their argument is an unconventional one. Essentially, they call for a focus on step two—the reasonableness step—of the inquiry when most courts have spent most of their time on step one. Perhaps partly for this reason, Kenneth Bamberger and Peter Strauss have tried to rehabilitate the *Chevron* two-step. They argue that the *Chevron* test “separates questions of statutory implementation assigned to independent judicial judgment (Step One) from questions regarding which the courts’ role is limited to oversight of agency decisionmaking (Step Two).”⁷⁴

This view recognizes that the two steps of the *Chevron* inquiry require two separate interpretive tasks and that the judiciary and the legislature are charged with two different roles in statutory interpretation. The first step implements the judicial role, a “pure” exercise in statutory interpretation, and the second step—a more untrammelled and deferential reasonableness inquiry—recognizes that Congress gave agencies the responsibility to implement its statutory instructions.

This approach is seemingly plausible, but it is practically divorced from the actual analysis with which courts are tasked. The idea that the first step of *Chevron* encourages courts to focus on statutory interpretation as a matter of judicial responsibility may be a useful pointer. But it has nothing to do with the basis on which the courts must evaluate the agency statutory interpretation. Courts do not evaluate the reasonableness of a particular statutory construction in a vacuum. Rather, they are, in the first step, using their own statutory interpretation tools to evaluate an *agency’s* interpretation. So characterizing *Chevron’s* first step as a call for courts to make a statutory interpretation finding before turning to the reasonableness of the agency’s interpretation, as Bamberger and Strauss do, adds a separation-of-functions curlicue to an analysis that would be difficult for courts to actually apply.

Stephenson and Vermeule are heirs to a line of scholars who have wondered whether even *Chevron’s* apparently straightforward—and certainly popular—test needlessly complicates an analysis of agency statutory interpretation that has long entailed

⁷⁴ Kenneth A. Bamberger & Peter L. Strauss. *Chevron’s Two Steps*, 95 Va. L. Rev. 611, 611 (2009).

some deference, but only deference apparently centered on reasonableness. Clark Byse believed that the *Chevron* two-step could be collapsed into a rule providing that courts ought to respect reasonable agency interpretations of their statutes.⁷⁵ Judge Williams has suggested that he has applied the inquiry as a single intellectual process.⁷⁶

Moreover, some recent practice suggests that courts are beginning to recognize that the reasonableness of agency action in a *Chevron* inquiry is critical. As Stephenson and Vermeule note, the Supreme Court announced in *Entergy Corp. v. Riverkeeper* that the agency “view governs if it is a reasonable interpretation of the statute.”⁷⁷ In *Global Crossing v. FCC*, the Court concluded, without paying much mind to *Chevron* at all, that the agency’s interpretation was “reasonable . . . [and] hence . . . lawful.”⁷⁸ If the Supreme Court is conflating *Chevron* steps one and two, it is worth thinking about how seriously other courts should take the process.

In theory, in cases where *Chevron* does not apply, courts are not required to defer to agency interpretations, regardless of their reasonableness; this is the *Mead* test, or “*Chevron* step zero,” which limits the applicability of *Chevron* to those instances where agencies are acting with force of law. *Mead* is, as I have noted, controversial. Many judges believe that adding a circumstance-dependent

⁷⁵ Clark Byse, *Judicial Review of Administrative Interpretation of Statutes: An Analysis of Chevron’s Step Two*, 2 Admin. L.J. 255, 256 n.10 (1988) (suggesting that the *Chevron* two-step test can be collapsed into a rule that courts must respect an agency’s reasonable interpretation of its organic statute).

⁷⁶ Erika Jones et al., *Developments in Judicial Review with Emphasis on the Concepts of Standing and Deference to the Agency*, 4 Admin. L.J. 113, 123–24, 126 (1990) (remarks of the Hon. Stephen F. Williams).

⁷⁷ 129 S. Ct. 1498, 1505 (2009). The Court noted in a footnote that:

The dissent finds it “puzzling” that we invoke this proposition (that a reasonable agency interpretation prevails) at the “outset,” omitting the supposedly prior inquiry of “whether Congress has directly spoken to the precise question at issue.” But surely if Congress has directly spoken to an issue then any agency interpretation contradicting what Congress has said would be unreasonable.

129 S. Ct. at 1505 n.4 (citation omitted).

⁷⁸ *Global Crossing Telecomms. v. Metrophones Telecomms.*, 550 U.S. 45, 47–48 (2007).

pre-test to the *Chevron* framework risks turning the test into something chaotic and unpredictable.⁷⁹

The *Mead* framework may not be a model of logical clarity, but perhaps some perspective is needed before it is deemed to be indecipherable by courts. If we believe that courts faced with confusing doctrine tend to resort to first principles, then it may be that courts take a more basic approach to review than the doctrine would suggest. In administrative law, those principles are best expressed as an agency reasonableness review, and there is good reason to believe that this baseline informs *Mead* review just as it informs the *Chevron* inquiry.

It is unlikely that a court would reverse a reasonable agency interpretation of a statute even if it were not obligated to show it deference under *Chevron* step two. *Chevron* step two requires courts to defer to reasonable agency legal interpretations; *Mead/Skidmore* instead requires courts to evaluate the reasonableness of an agency determination *without* requiring them to defer to it. It is this distinction of *Mead* that suggests that courts may not treat *Chevron* and *Mead* cases so very differently. Because they are obligated under *Mead* to consider whether the agency interpretation is reasonable (but then are not obligated to defer to it), it is only in those cases where a bold court decides that an agency interpretation, albeit reasonable, ought not be preferred where there is any gap in the review, at least if one takes the doctrine seriously. And it is by no means clear that courts would act in such a manner with any frequency.

Moreover, the cases and scholars applying *Mead* suggest as much. And indeed, the Ninth Circuit too has suggested in a fishing permit case that the *Chevron* analysis would be no different from the *Skidmore* analysis in precisely the sort of complex environ-

⁷⁹ Dissenting from the *Mead* decision itself, Justice Scalia said making some agency actions in furtherance of their regulatory mission subject to strong deference, while others would be subject to more limited deference, is “neither sound in principle nor sustainable in practice.” *United States v. Mead Corp.*, 533 U.S. 218, 241 (2001) (Scalia, J., dissenting). The lower courts have also expressed some displeasure. See *supra* note 64 and accompanying text. For an introduction to the academic literature, see Lisa Schultz Bressman, *How Mead Has Muddled Judicial Review of Agency Action*, 58 *Vand. L. Rev.* 1443, 1445 (2005).

mental case where one might think there would be a difference.⁸⁰ In a water pollution dispute, the Ninth Circuit concluded that “[i]n the end, though, it does not much matter . . . whether we review the EPA’s position through the *Chevron* or *Skidmore/Mead* prism.”⁸¹ Other circuit courts have done the same thing, leading Amy Wildermuth to observe that there are more than a few cases “in which courts refuse to decide whether they are affording the agency position *Chevron* or *Skidmore* deference.”⁸² Even critics of the *Mead* doctrine acknowledge that in many cases the putatively different tests represented by *Mead* and *Chevron* would be unlikely to lead to different outcomes.⁸³

Furthermore, the animating hand of reasonableness, apparent in both the *Mead* and *Chevron* standards of review, is just as difficult to remove from de novo review. In de novo review, of course, the court is not required to consider the agency decision at all. And although de novo review does not lie at the heart of administrative law—it is not particularly common, though it is a standard that applies whenever agencies interpret the APA and the Constitution—

⁸⁰ *Wilderness Soc’y v. U.S. Fish & Wildlife Serv.*, 316 F.3d 913, 922 (9th Cir. 2003) (concluding, after a long discussion admitting the complexity of the interaction between the two standards, that *Chevron* applies but that the result is the same as it would be under *Skidmore*).

⁸¹ *Pronsolino v. Nastri*, 291 F.3d 1123, 1135 (9th Cir. 2002).

⁸² Amy J. Wildermuth, *Solving the Puzzle of Mead and Christensen: What Would Justice Stevens Do?*, 74 *Fordham L. Rev.* 1877, 1895–96 (2006). For examples, as Wildermuth notes, see *Pension Benefit Guar. Corp. v. Wilson N. Jones Mem’l Hosp.*, 374 F.3d 362, 369 (5th Cir. 2004) (“We do not need to decide whether the [order] . . . warrants *Chevron* deference because it is clear that the . . . order may be upheld as a matter of law under the less deferential standard set forth in [*Mead*].”) and *Cnty. Health Ctr. v. Wilson-Coker*, 311 F.3d 132, 138 (2d Cir. 2002) (affording an interpretation “considerable deference, whether under *Chevron* or otherwise”) and compare *United States v. Atandi*, 376 F.3d 1186, 1189 (10th Cir. 2004) (“Without determining whether full *Chevron* deference is owed . . . we unquestionably owe ‘some deference’ to the . . . regulation.” (citation omitted)).

⁸³ See, e.g., Adrian Vermeule, Introduction: *Mead* in the Trenches, Recent Decisions of the United States Court of Appeals for the District of Columbia Circuit, 71 *Geo. Wash. L. Rev.* 347, 350–51 (2003) (“The costs of the elaborate predecision required by *Mead* will be highest whenever the difference between *Chevron* deference and *Skidmore* deference will make *no difference* to the resolution of the ultimate statutory question.” (emphasis added)). This is not to say that there is no meaningful difference in the choice of test—parties will incur significantly higher costs under the *Mead* regime—but simply that a given case’s ultimate outcome may well be unaffected by that choice.

it is just as susceptible to the structural inclinations towards reasonableness.

This tendency results from the requirement that courts must always, as a first step, assess the soundness of an agency decision. Of course, courts are then—at least in theory—entitled to reject the agency decision and draw their own conclusions. As we will see, the empirical evidence here is intriguing, although not dispositive. But courts cannot review agency legal interpretations *de novo* simply by disregarding those interpretations. Indeed, the doctrines of ripeness, mootness, timing, and jurisdiction require them to address those issues before proceeding to the next part of the case.⁸⁴

In other words, courts have to consider the agencies' positions before coming up with their own view even in *de novo* review cases. And while there is reason to think that the review of the agency's case might “anchor” the court into considering the merits of it before engaging in its own legal interpretation, human nature alone need not serve as the only basis of such a conclusion. After all, in habeas review, some courts have concluded that “the distinction between *de novo* review and ‘reasonableness’ review becomes insignificant” given the standard of review that Congress has given them.⁸⁵

So while in theory, *de novo* review is a very different standard from that of reasonableness, in practice it is difficult to see how courts would be able to ignore reasonable agency interpretations in reaching their conclusions.⁸⁶ The Supreme Court's experience with *de novo* review, moreover, only bolsters this observation, as we will see.

⁸⁴ The Supreme Court has long held that these jurisdictional issues could not be disregarded with an eye to the merits, though it has occasionally been posited that this dictum is sometimes observed in the breach. *Mansfield, Coldwater & Lake Mich. Ry. Co. v. Swan*, 111 U.S. 379, 382 (1884) (“[T]he first and fundamental question is that of jurisdiction This question the court is bound to ask and answer for itself, even when not otherwise suggested . . .”).

⁸⁵ *Cardwell v. Greene*, 152 F.3d 331, 338 (4th Cir. 1998) (quoting *Cardwell v. Netherland*, 971 F.Supp 997, 1015 (E.D. Va. 1997)), overruled by *Bell v. Jarvis*, 236 F.3d 149 (4th Cir. 2000).

⁸⁶ Judge Richard Posner might disagree. As he has explained, “the cognitive limitations that judges share with other mortals may constitute an insuperable obstacle to making distinctions any finer than that of plenary versus deferential review.” *School Dist. of Wis. Dells v. Z.S. ex rel. Littlegeorge*, 295 F.3d 671, 674 (7th Cir. 2002).

B. Arbitrariness Review Is Inseparable from the Reasonableness Inquiry that Is the Basis for Review of Law

Students and lawyers are told that *Chevron* review of legal interpretations by agencies and “hard look” review of their final decisions for policymaking rationality are two different tests, both of which must be applied to any case raising legal interpretation issues.⁸⁷ The idea is that a court must first decide whether the agency is interpreting its legal authority reasonably, and then secondly, that it should decide whether the action is, in a more general sense, arbitrary or capricious.

But we do not need two tests for these sorts of cases. If *Chevron* review does in fact turn on reasonableness, the possibility that an agency interpretation of law might be presumed to be reasonable despite failing the arbitrary and capricious test is awfully unlikely. Indeed, to my knowledge, no court has concluded that reasonable legal interpretations are nonetheless components of arbitrary agency action.

It is accordingly unsurprising that, in cases like *Arent v. Shalala*, the D.C. Circuit has acknowledged an “overlap” between *Chevron* and arbitrariness review.⁸⁸ In *National Association of Regulatory Utility Commissioners v. Interstate Commerce Commission*, the court noted that it was possible to conclude that a reasonableness inquiry might lead to the same outcome “whether one considers the case as one involving a question of *Chevron* Step II statutory interpretation or a garden variety arbitrary and capricious review or, as we do, a case that overlaps both administrative law concepts.”⁸⁹ More generally, courts have cited *State Farm* cases when performing a *Chevron* step two analysis,⁹⁰ and vice versa.⁹¹

⁸⁷ See supra notes 1–3 and accompanying text.

⁸⁸ *Arent v. Shalala*, 70 F.3d 610, 616 n.6 (D.C. Cir. 1995) (“The *Chevron* analysis and the ‘arbitrary, capricious’ inquiry set forth in *State Farm* overlap in some circumstances, because whether an agency action is ‘manifestly contrary to the statute’ is important both under *Chevron* and under *State Farm*.”); id. at 620 (Wald, J., concurring) (“I agree with the panel that despite these distinctions, the *Chevron* and *State Farm* frameworks often do overlap.”).

⁸⁹ 41 F.3d 721, 728 (D.C. Cir. 1994).

⁹⁰ See, e.g., *Am. Fed’n of Labor & Cong. of Indus. Orgs. v. Brock*, 835 F.2d 912, 917 (D.C. Cir. 1987).

⁹¹ A Guide to Judicial and Political Review of Federal Agencies 96, 180 (John F. Duffy & Michael Herz eds., 2005) [hereinafter ABA Judicial Review Guide]

What should we make of this judicial gloss? As Ronald Levin has said, we can probably rest assured that, at least before the D.C. Circuit, there is little meaningful difference between *Chevron* step two and *State Farm*, for “the D.C. Circuit has often drawn upon the methods of arbitrariness review in its elaboration of *Chevron* step two.”⁹²

The judiciary’s growing unwillingness to distinguish between *Chevron* step two and arbitrariness review has led a number of observers to urge that the two standards be conflated. The American Bar Association’s (ABA) Administrative Law Section has urged that “*Chevron* step two should be explicitly understood to incorporate the requirements of arbitrary and capricious review” given that “the arguments with respect to both may often be similar,” and “distinguishing conceptually between” the two standards is “difficult.”⁹³

Stephenson and Vermeule go even further down the “one standard to rule them all” road when comparing *Chevron* and *State Farm*. In their view, “it takes no more intellectual work, and no more disruption of existing doctrine, to collapse Steps One and Two,” which is their insight, “than it does to collapse Step Two and *State Farm*.”⁹⁴ They thus intimate that between two of the most cited standards of review in administrative law—*Chevron* and *State Farm*—standards that are supposed to be kept distinct by the courts—there is little difference, and a shared focus on reasonableness.

In this, they are joined by Levin and the scholars who participated in the ABA’s guide to administrative law. The scholarly consensus is, I think, warranted; it is difficult to imagine, from a purely

(“[M]any cases applying the arbitrary and capricious test that are discussed in this chapter are assessing the ‘reasonableness’ of an agency interpretation as part of the *Chevron* step two inquiry.”).

⁹² Ronald M. Levin, *The Anatomy of Chevron: Step Two Reconsidered*, 72 *Chi.-Kent L. Rev.* 1253, 1271 (1997). Levin notes that the court “has stopped short of actually calling them identical,” and argues “that this insistence on a possible divergence is misdirected.” *Id.*

⁹³ ABA Judicial Review Guide, *supra* note 91 at 96, 100.

⁹⁴ Stephenson & Vermeule, *supra* note 70, at 604 n.28; cf. Jerry L. Mashaw, *Reasoned Administration: The European Union, the United States, and the Project of Democratic Governance*, 76 *Geo. Wash. L. Rev.* 99, 111 (2007) (noting that administrative law has been structured to permit “judicial review of policy choices for reasonableness”).

textualist perspective, that appellate judges consider reasonableness to be a very different concept from a lack of arbitrariness. Would it ever be the case that a reasonable interpretation of an agency's legal authority, as applied to a specific set of facts, would lead it to an arbitrary result? It is difficult to conceptualize such a case, and harder still to find an example of such a case in the pages of the Federal Reporter. Likewise, it is hard to imagine that an agency's arbitrary action would nonetheless be based on a reasonable understanding of its statutory authorization.

While the distinction between *Chevron* and *State Farm* is plainly elusive in actual court decisions, the doctrinally inclined might wonder whether the analogy could hold true for the other legal standards of review, namely *Mead/Skidmore* and de novo review. Can these standards also be conflated with the arbitrariness inquiry?

If anything, the logical case for conflation is even easier to argue. Those who may believe (as I do not) that courts actually treat the proposed legal basis for agency action less deferentially in *Skidmore/Mead* and de novo cases than they do in *Chevron* cases will be hard-pressed to come up with an example where a court, with no *obligation* to defer, nonetheless did so, but still found the agency action to be arbitrary or capricious. I am unaware of such a case, and it is difficult to imagine one.⁹⁵ It is accordingly all but impossible to imagine an agency action surviving any of the tests for legal authority without also surviving the arbitrariness test.

In fact, it may be better to think of the arbitrariness test as something without much substantive bite at all. Rather, *State Farm* (and its antecedent, *Overton Park*) is best understood as chiefly imposing procedural obligations on agencies, so that courts can act in the few cases where they feel comfortable concluding that an agency action is completely beyond the pale, which they can only do if they have an adequate explanation of what, exactly, the agency was

⁹⁵ Imagination would have to turn to something more suited to the professorial hypothetical than the real world, such as where an agency might interpret the comment requirement of the APA to permit comment, but only in a very particular format somehow related to the agency's mission—with balance sheets for SEC accounting rules, say, or in Spanish for particular aspects of DHS enforcement of the immigration laws.

doing.⁹⁶ The arbitrariness requirement has been interpreted by many courts and commentators as a “giving reasons” requirement, which does sound procedural, and so its bite lies in its requirement that agencies explain themselves.⁹⁷ *State Farm* itself has language that reflects the importance of this procedural requirement: the agency “must examine the relevant data and articulate a satisfactory explanation for its action including a ‘rational connection between the facts found and the choice made.’”⁹⁸ The *Overton Park* decision that preceded *State Farm* similarly required courts to conduct a “searching and careful” inquiry to determine “whether the [agency’s] decision was based on a consideration of the relevant factors.”⁹⁹

What should we make of the procedural component of arbitrariness review? In my view, it makes that review a procedural variant of the more general reasonableness requirement on agency action. “Giving reasons” is very difficult to distinguish from acting reasonably and not just as a matter of semantics. Reasonable agency actions *are* explained and justified, and reasonable legal interpretations are not simply left to the courts to evaluate *sans* guidance. Rather, those conclusions of law are spelled out in Federal Register notices and appellate briefs. These are the tools that agencies use to give reasons for their actions. And in so doing, they both address the legal authority they have to act and, at the same time, meet the requirements of the *State Farm* test. Meanwhile, failing to provide an explanation of what an agency is doing will be seen

⁹⁶ The substantive aspect of *State Farm* review also, one suspects, serves as a failsafe to keep courts from immersing themselves in procedural technicalities, such as determining whether an agency has really met all of its notice and comment obligations when the result of that rulemaking seems bizarre.

⁹⁷ For leading discussions of this “explain yourself” view, see Martin Shapiro, *The Giving Reasons Requirement*, 1992 U. Chi. Legal F. 179 (outlining this requirement as a matter of comparative administrative law), and Mashaw, *supra* note 94, at 101 (seeking, in a comparison of the U.S. and European Union systems, to “explain why reason giving is so prominent a part of both administrative systems, how it functions juridically, and, most crucially, what the reasons are for demanding reasons or for providing a ‘right’ to reasoned administration”); see also Frederick Schauer, *Giving Reasons*, 47 Stan. L. Rev. 633, 634 (1995) (discussing the requirement as a component of judicial, as well as agency, decisionmaking).

⁹⁸ *Motor Vehicle Mfrs. Ass’n of the U.S. v. State Farm Mut. Auto. Ins. Co.*, 463 U.S. 29, 43 (1983) (quoting *Burlington Truck Lines v. United States*, 371 U.S. 156, 168 (1962)).

⁹⁹ *Citizens to Pres. Overton Park, Inc. v. Volpe*, 401 U.S. 402, 416 (1971).

as unreasonable both when the arbitrariness of a decision is at issue and in *Chevron* step two.

C. Arbitrariness Review and Fact Review Are Identical Standards

Agency factfinding, as I have noted, is, at least in theory, subject to different standards of review, depending on whether formal or informal proceedings are involved. But in the real world of judicial review, the lines between the two tests are far from clear. In *Overton Park*, the Supreme Court held that review of informal agency action must be “thorough, probing, in-depth” as well as “searching and careful”—even where the substantial evidence test does not apply.¹⁰⁰ In *Universal Camera*, the Court also provided that substantial evidence review be subject to this sort of scrutiny, noting, for example, that “[t]he substantiality of evidence must take into account whatever in the record fairly detracts from its weight.”¹⁰¹ Moreover, the *Universal Camera* decision that defined substantial evidence review is replete with paeans to reasonableness. It concluded that “courts must now assume more responsibility for the *reasonableness* and fairness of Labor Board decisions,” partly by looking to the factual findings of the administrative judge who took the evidence, and giving them the “relevance that they *reasonably* command in answering the comprehensive question whether the evidence supporting the Board’s order is substantial.”¹⁰²

Accordingly, lower courts have increasingly wondered why the Supreme Court bothers to distinguish between substantial evidence and factfinding arbitrariness review. In *Association of Data Processing Service Organizations v. Board of Governors of the Federal Reserve System*, the D.C. Circuit explained, in an opinion by then-Judge Scalia and joined by then-Judge Ginsburg, that “in their application to the requirement of factual support the substantial evidence test and the arbitrary or capricious test are one and the same.”¹⁰³ Logically, the lower court reasoned, “it is impossible to conceive of a ‘nonarbitrary’ factual judgment supported only by

¹⁰⁰ Id. at 415–16; see also *Associated Indus. of N.Y. State v. U.S. Dep’t of Labor*, 487 F.2d 342, 349–50 (2d Cir. 1973) (identifying convergence between “substantial evidence” and “arbitrary or capricious” criteria).

¹⁰¹ *Universal Camera Corp. v. NLRB*, 340 U.S. 474, 488 (1951).

¹⁰² Id. at 490, 497 (emphasis added).

¹⁰³ 745 F.2d 677, 683 (D.C. Cir. 1984).

evidence that is not substantial.”¹⁰⁴ Since then, the D.C. Circuit has indicated that, in its view, substantial evidence review and arbitrary and capricious review “connote[] the same substantive standard[s] of review.”¹⁰⁵

The professoriate is in agreement with the judiciary. The ABA’s Administrative Law Section has suggested that “it is difficult to perceive any reason in principle why the distinction [between substantial evidence factfinding and arbitrary factfinding] should be preserved.”¹⁰⁶ Richard Pierce, in his *Administrative Law Treatise*, has concluded that “courts frequently treat the two tests as identical, referring to their ‘tendency to converge’ and to the distinction between the two as ‘largely semantic,’” because “[w]hatever the arbitrary and capricious test may require, it cannot demand more evidentiary support than the substantial evidence test.”¹⁰⁷

Although the Supreme Court has not issued a definitive pronouncement on the question, it is fair to say that most observers, most appellate judges, and binding precedent in the D.C. Circuit all have found that the two standards of review for agency factfinding may be conflated into one.¹⁰⁸

But does this mean that the one standard of review of agency factfinding is a reasonableness standard, not worth distinguishing from the reasonableness inquiries undertaken in the other standards of review? There is every reason to conclude that it means exactly that. After all, agency factfinding and arbitrariness review are based on the *exact same* provision of the APA, Section 706(2)(A). It has never been easy to understand why *State Farm* review would be different from informal factfinding review, given that their legal basis stems from the same statutory subsection.

¹⁰⁴ Id. at 684.

¹⁰⁵ *Bangor Hydro-Elec. Co. v. Fed. Energy Regulatory Comm’n*, 78 F.3d 659, 663 n.3 (D.C. Cir. 1996).

¹⁰⁶ ABA Judicial Review Guide, *supra* note 91, at 162.

¹⁰⁷ 2 Richard J. Pierce Jr., *Administrative Law Treatise* 806, 807 (4th ed. 2002).

¹⁰⁸ For a recent Supreme Court decision that obscures the distinction between the sorts of tests involved in fact cases that also implicate *Chevron*, see *United States v. Eurodif S.A.*, 129 S. Ct. 878, 886–87 & n.7 (2009). For a critique on its choice of standard of review standard, see Posting of Jonathan Adler to The Volokh Conspiracy, Funk on *Eurodif & Chevron*, <http://www.volokh.com/posts/1233170920.shtml> (Jan. 28, 2009, 2:28 pm).

Courts have not suggested that their approach varies depending on the type of the agency action under review.

Even academics do not distinguish between questions of fact and question of law when they are evaluating administrative law. For example, Thomas Miles and Cass Sunstein studied arbitrariness review by looking at the NLRB, whose factfinding is reviewed under the substantial evidence test, and the EPA, whose factfinding is usually not.¹⁰⁹ They concluded that regardless, the two agencies' outcomes in the courts of appeals could usefully be compared; "[i]n 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," they posited.¹¹⁰ Indeed, as they note, the legislative history of the statute at issue in *State Farm*, which announced the definitive arbitrariness test, called for substantial evidence review of the agency's findings.¹¹¹ For Miles and Sunstein, the Court seemed to suggest that the substantial evidence test—even though it involved review of fact, rather than more general arbitrariness review—was “essentially identical” to *State Farm* review.¹¹²

In sum, the two factfinding standards have been conflated by observers, and the analogy between those standards and *State Farm* review is hard to gainsay, given that they both stem from the same statutory authorization. Accordingly, what initially looks like six different standards of review—and what has been announced by the courts as such—does not, at bottom, look so very different at all. Instead, faced with a set of elaborate doctrinal requirements, the courts have simplified matters. They have looked to the reasonableness of the agency's action, whatever the sort of action it was. They have analogized one standard of review to another. They, along with scholars, have expressed their doubts as to whether the legal analysis would really differ, regardless of which standard would apply. And so, even doctrinally, it is quite clear that it is the reasonable agency standard that applies to almost every case of administrative law.

¹⁰⁹ Thomas J. Miles & Cass R. Sunstein, *The Real World of Arbitrariness Review*, 75 U. Chi. L. Rev. 761 (2008) [hereinafter Miles & Sunstein, *Arbitrariness*].

¹¹⁰ *Id.* at 764.

¹¹¹ *Id.* at n.25.

¹¹² *Id.*

III. STANDARDS OF REVIEW IN PRACTICE

Realists might view the doctrinal analogizing in the prior Section as something of an arid exercise, and, to be sure, the recommendations in this Article do not depend on the logical consistencies—or inconsistencies—among the standards of review, as the courts have spelled them out. It is at least as interesting to sort out what courts actually *do* when they review agency action, be it of fact, of law, or of policymaking. Here, the evidence is striking. Courts reverse agencies at roughly the same rate, regardless of the standard of review. A study of those decisions suggests that the win rates of agencies are surprisingly consistent across standards of review. For *Chevron* review, *Skidmore* review, and arbitrary and capricious review, the studies suggest that, at least as the judiciary is currently comprised, agencies win between 60 and 70% of their appeals with few exceptions.

Even the more rarely invoked standards, such as the *Skidmore* review of legal determinations or *de novo* review, appear to fall within the basic zone of judicial affirmation rates. This Article adds substantial evidence review to these studies; there, too, agencies win 70% of their appeals. The results are within a bandwidth of commonality suggesting that, regardless of the doctrinal and contextual differences among the six standards, the ensuing judicial review is the same. Although there are a number of possible conclusions to draw, the doctrinal one worth taking most seriously is that, unless there is some reason to believe that these very similar validation rates mask very different sorts of inquiries,¹¹³ what courts are really doing is the same sort of analysis regardless of the standard of review. The consistency in outcomes suggest a consistent inquiry: courts look to see if the agency has acted reasonably.

The point of going through these win-rate exercises is to suggest that scholars and judges, in doubting that the standards of review are very different as a logical matter, are not far off the mark: the standards are, with little doubt, analogous to each other in actual practice. In their decisionmaking, courts pick winners at the same rate regardless of the standard of review, which suggests that they

¹¹³ And I have no cause to think that there are such reasons—nor does the doctrine, given the difficulties that exist between separating the borders of many of the judicial review tests from one another.

do not appear to be making categorical distinctions among fact cases, law cases, and arbitrariness cases.¹¹⁴

A. Prior Scholarship on Agency Validation

Prior scholarship on the validation of agencies reveals a surprisingly uniform rate at which agencies win their cases. There is a vibrant and growing set of studies on how agencies fare in various judicial review contexts. The literature, however, lacks an article that puts these empirical insights together. I do so here, and find consistencies where pure doctrine would lead lawyers to believe there might be differences. The point is best proved by pooling the studies and by weighting them according to the number of cases involved. The assumption is that, even if one study is unreliable, averaging a number of studies is more likely to approach something like the “market price” of appellate validation of agency action, regardless of the standard of review.¹¹⁵ The dataset includes eleven studies consisting of 5081 observations across all judicial review standards. The starting point of my literature review is 1982, corresponding to the post-*Chevron* period. When the eleven studies are pooled, the overall agency validation rate is 69%. This rate is remarkably consistent with the validation rates each of these studies finds for the various inquiries they conducted, across issue areas, standards of review, and time. The low variance makes a strong case for an amalgamated standard of review.

¹¹⁴ Selection bias is an issue with these sorts of comparisons, of course—different litigants may bring different sorts of cases if they are confronted by different standards of review. But the fact that agency validation differs little among all of the standards of review offers some comfort to the claim here.

¹¹⁵ See Bebachuk et al., *supra* note 20, at 1790–92.

2010]

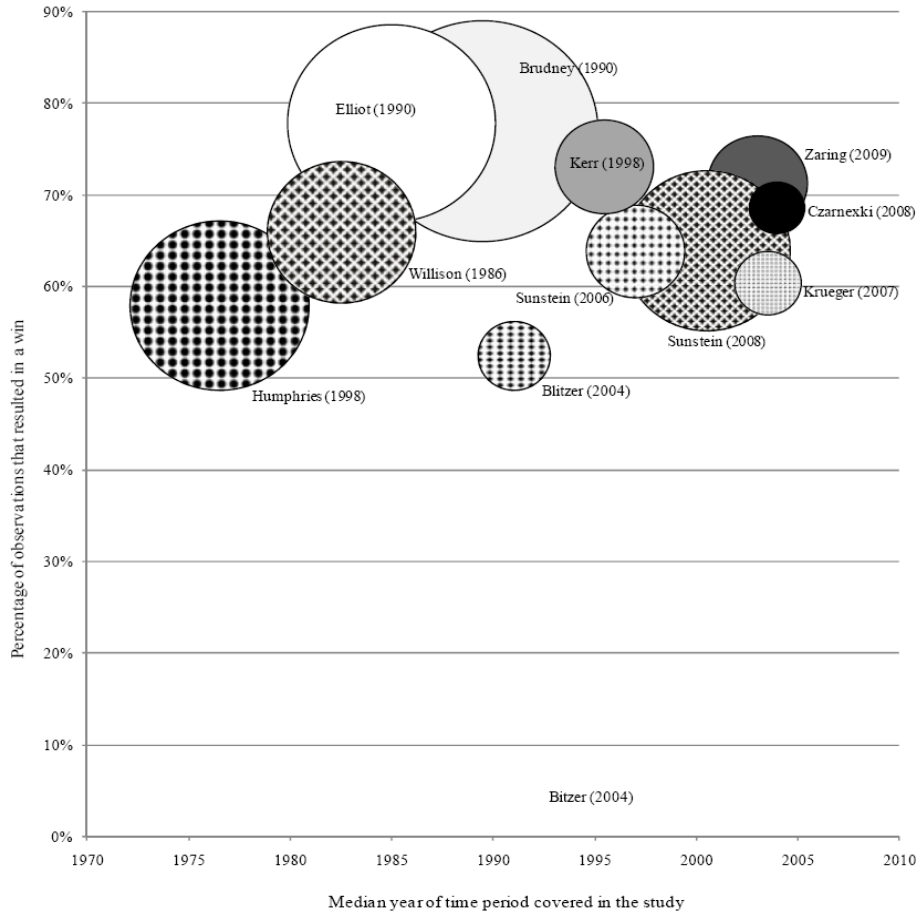
Reasonable Agencies

171

Table 1: Agency Validation Studies (% reflects percentage of decisions in which agency won)

Authors	Year	Study	Time period	Wins	Observations	%
Zaring	2009	D.C. Circuit, substantial evidence	2000–2004	161	226	71
Miles Sunstein	2008	Courts of appeals, A&C or sub. ev.	1995–2006	418	653	64
Czarnezki	2008	Environmental cases, <i>Chevron</i>	2003–2005	48	70	69
Hickman Krueger	2007	Courts of appeals	2001–2006	64	106	60
Miles Sunstein	2006	Courts of appeals, <i>Chevron</i>	1990–2004	145	227	64
Caruson Bitzer	2004	D.C. Circuit	1986–1996	64	122	54
Brudney, et al.	1999	Courts of appeals, unfair labor practice	1986–1993	944	1228	76
Kerr	1998	Courts of appeals, <i>Chevron</i>	1995–1996	163	223	73
Songer Humphries	1998	Courts of appeals	1959–1994	426	734	58
Shuck Elliot	1990	Courts of appeals, <i>Chevron</i>	1985	763	981	77
Willison	1986	D.C. Circuit	1981–1984	337	511	66
Total				3533	5081	69

Figure 1: Agency Validation Studies by Year Studied (size of study reflects number of observations).¹¹⁶



¹¹⁶ Thanks to Christina Yi for assistance with the figure.

A number of authors have conducted empirical studies on this topic; I collected all such studies cited in the legal and political science literature. The overarching studies revealed a validation rate quite close to the two-thirds validation rate posited here, with the exception of Kiki Caruson's early study of the D.C. Circuit (when liberal judges roared) and the validation rate for agency action was approximately 54%.¹¹⁷ No other study matched this result, although Martha Humphries and Donald Songer's study, which included both pre- and post-*Chevron* cases, yielded a validation rate in the courts of appeals of 58% between 1969 and 1988.¹¹⁸ James Brudney's co-authored study of appellate validation of the NLRB in unfair labor practices, however, gleaned a significantly higher validation rate of 76%, regardless of the standard used.¹¹⁹ That study, although aging, reviewed all appellate decisions in these sorts of cases between 1986 and 1993.¹²⁰ Not included in this study is Frank Cross' analysis of the cases in the Songer database; it is not clear exactly how many of these cases (which are a sample of appellate cases) apply to administrative law, and the coding is somewhat imprecise. Nevertheless, Cross' extraction of data from the database is broadly consistent with this study; he reports an affirmance rate of 67% for de novo review, 70% for substantial evidence, and 75% for arbitrary and capricious review.¹²¹

Chevron cases, because they explicitly include a reasonableness requirement (and because *Chevron* is the most cited case in administrative law) might serve as a baseline. How do agencies fare when courts are reviewing their interpretations for reasonableness? In an early study, Donald Elliott and Peter Schuck identified a rather high validation rate of 77% in *Chevron*'s immediate aftermath.¹²² Miles and Sunstein found that circuit judges voted to uphold

¹¹⁷ Kiki Caruson & J. Michael Bitzer, At the Crossroads of Policymaking: Executive Politics, Administrative Action, and Judicial Deference by the DC Circuit Court of Appeals (1985–1996), 26 L. & Pol'y 347, 360 (2004).

¹¹⁸ Martha Anne Humphries & Donald R. Songer, Law and Politics in Judicial Oversight of Federal Administrative Agencies, 61 J. Pol. 207, 215 (1999).

¹¹⁹ James J. Brudney, Sara Schiavoni & Deborah J. Merritt, Judicial Hostility Toward Labor Unions? Applying the Social Background Model to a Celebrated Concern, 60 Ohio St. L.J. 1675, 1714 (1999).

¹²⁰ Id. at 1694.

¹²¹ Frank B. Cross, Decision Making in the U.S. Courts of Appeals 53 (2007).

¹²² Peter H. Schuck & E. Donald Elliott, To the *Chevron* Station: An Empirical Study of Federal Administrative Law, 1990 Duke L.J. 984, 1007–09 (1990).

agency action in *Chevron* cases approximately 64% of the time.¹²³ Other *Chevron* studies are similar. Jason Czarnezski examined environmental cases between 2003 and 2005 in the courts of appeals and found agency decisions that were reviewed under the *Chevron* standard were affirmed 69.2% of the time.¹²⁴ In a somewhat older study, Orin Kerr found that between 1995 and 1996, agencies won 73% of their *Chevron* cases. As he said then, “[i]n the years 1995 and 1996, the U.S. Courts of Appeals applied the *Chevron* doctrine in 223 published cases that fit the criteria of this study. In these 223 cases, the courts applied the doctrine 253 times. The agency interpretation was accepted in 73% of these applications.”¹²⁵ These studies suggest some variance, but not too much: agencies have won their *Chevron* cases roughly two thirds of the time in the courts of appeals regardless of the agency—even if the time period in question varies by a decade or so.

What about those cases where agencies make legal interpretations, but where *Chevron* does not apply? As I have noted, those are the cases where courts are not *required* to defer to reasonable agency action, though they may do so if they wish. So there is a doctrinal reason to suspect—even if it is not a very good one¹²⁶—that those cases would be treated differently.

But agency legal interpretations do not fare worse when *Chevron* does not apply, at least not dramatically so. Instead, the outcomes are consistent. Kristin Hickman and Matthew Krueger examined 106 *Skidmore* cases and found that the courts accepted the

¹²³ Thomas J. Miles & Cass R. Sunstein, Do Judges Make Regulatory Policy? An Empirical Investigation of *Chevron*, 73 U. Chi. L. Rev. 823, 849 (2006) [hereinafter Miles & Sunstein, *Chevron*]. The validation rate for Supreme Court justices, in aggregate, was approximately 67%, *id.*, though some justices, such as Breyer and Souter, voted for the agency 79% of the time, while Scalia and Thomas voted for the agency approximately 53% of the time. *Id.* at 831.

¹²⁴ Jason J. Czarnezski, An Empirical Investigation of Judicial Decisionmaking, Statutory Interpretation, and the *Chevron* Doctrine in Environmental Law, 79 U. Colo. L. Rev. 767, 782–83 (2008).

¹²⁵ Orin S. Kerr, Shedding Light on *Chevron*: an Empirical Study of the *Chevron* Doctrine in the U.S. Courts of Appeals, 15 Yale J. on Reg. 1, 30 (1998). Like Kerr’s mid-nineties study, earlier studies of judicial review are not quite as uniform as this paper suggests. See Schuck & Elliott, *supra* note 122, at 1029–31 (analyzing 2000 decisions by the courts of appeals that document a pre-*Chevron* affirmance rate of 71% versus a post-*Chevron* rate of 81%).

¹²⁶ See *supra* Section II.A.

agency's interpretation in 64 of the 106 cases (60.4%) and that the Court rejected the agency's interpretation in 42 of the 106 cases (39.6%).¹²⁷ The rate is lower than *Chevron*, but not notably so. Moreover, for those agency interpretations that are not accorded any deference at all—as, for example, when the agency is interpreting the Constitution, or statutes that they are not charged with administering, like Title VII or the APA itself—agencies do just as well, though data in this area are limited.¹²⁸ In de novo review cases before the Supreme Court, at least according to Eskridge and Baer's data, the agency win rate is 66%—again, very close to the *Chevron* validation rates reported by both Czarneski and Miles and Sunstein.¹²⁹

Arbitrary and capricious review under *State Farm's* so-called “hard look” standard is not very different. In another study, Miles and Sunstein found that, in a dataset consisting of two agencies, the agency won its *State Farm* review approximately 64% of the time, though the win rates of the two agencies they studied were different: EPA decisions were affirmed 72% of the time, compared to 62% for the NLRB.¹³⁰ (It may be that the agency promulgating the administrative action is a better predictor of whether the action will be affirmed than are the various standards of review, though the original empirical work in this paper does not support that conclusion.)

Of particular note here is the fact that two authors—Miles and Sunstein—did exactly the same sort of validation study on two different forms of judicial review, *and found precisely the same validation rate*. The surprising thing to me is that they did not trumpet

¹²⁷ See Kristin E. Hickman & Matthew D. Krueger, In Search of the Modern *Skidmore* Standard, 107 Colum. L. Rev. 1235, 1275 (2007).

¹²⁸ William N. Eskridge, Jr. & Lauren E. Baer, The Continuum of Deference: Supreme Court Treatment of Agency Statutory Interpretations from *Chevron* to *Hamdan*, 96 Geo. L.J. 1083, 1099, 1117–19 (2008) (finding a 66% agency win rate with no deference).

¹²⁹ *Id.* at 1090 (“Indeed, our most striking finding is that in the majority of cases—53.6% of them—the Court does not apply any deference regime at all. Instead, it relies on ad hoc judicial reasoning of the sort that typifies the Court’s methodology in regular statutory interpretation cases.”).

¹³⁰ Miles & Sunstein, *Arbitrariness*, *supra* note 109, at 776, 778–79.

this observation—although perhaps this Article is better placed to do so.¹³¹

These studies and observations are consistent with some of the literature's first win rate studies, which found similar win rates for agencies even before the doctrinal evolution of the standards we have today.¹³² They are also consistent with a relatively large academic literature that has concluded that win rates may stabilize around any probability of victory.¹³³

If courts affirm agencies in their review of law decisions 60 to 73% of the time, with a 64% average in *Chevron* cases, and if they affirm agencies in their arbitrariness decisions approximately 64% of the time, there appears to be something striking going on.¹³⁴

¹³¹ Miles and Sunstein were particularly concerned with the politicization of judicial review, both by party and by group dynamics. Although there is no question that ideology is an important predictor of many sorts of judicial review outcomes, this Article is principally occupied with the more fundamental questions: Around what baseline? And at what variance?

¹³² See Humphries & Songer, *supra* note 118, at 215 tbl.1; David H. Willison, *Judicial Review of Administrative Decisions: Agency Cases before the Court of Appeals for the District of Columbia, 1981–84*, 14 *Am. Pol. Q.* 317, 320–21 (1986).

¹³³ Which is one reason why the consistency of the two-thirds win rate need not detain us for too long. In a well-known article, George Priest and Benjamin Klein argued that experience would lead litigants to select cases for litigation that would lead plaintiffs to win approximately 50%. See George L. Priest & Benjamin Klein, *The Selection of Disputes for Litigation*, 13 *J. Legal Stud.* 1, 4 (1984) (hypothesizing that the only cases brought to trial would be those in which the outcome was uncertain, and in cases where the plaintiff or defendant was likely to win, settlement would likely occur). Empirically testing the hypothesis, however, did not confirm it. See Theodore Eisenberg, *Testing the Selection Effect: A New Theoretical Framework with Empirical Tests*, 19 *J. Legal Stud.* 337, 350 (1990). Scholars such as Steven Shavell identified reasons why the 50% rule would not hold. See Steven Shavell, *Any Frequency of Plaintiff Victory at Trial Is Possible*, 25 *J. Legal Stud.* 493, 494 & n.7 (1996) (discussing the limitations of the Priest and Klein model with regard to high value cases, low transaction cost cases, and the like).

¹³⁴ This Article only addresses the rates in the courts of appeals (though they are occasionally compared to the Supreme Court). District court review of agency action may present different circumstances; Paul Verkuil's research suggests as much, anyway. For Social Security Administration (SSA) appeals, "[t]he rate at which district courts reverse and remand disability determinations exceeds 50%." Verkuil, *supra* note 5, at 706. In contrast, Verkuil finds that the district court reversal rate for Freedom of Information Act (FOIA) cases is only 10%. *Id.* at 712–13. This is actually quite surprising, but it could be that the district courts, the experts at factual analysis, are more inclined to weigh in on appeals of fact-based agency determinations—which SSA determinations require—than legal interpretations, which, at least in theory, FOIA interpretations are. Verkuil's methods are aggregative, and do not involve a review of individual cases. *Id.* at 713 n.152.

B. Judicial Review of Agency Factfinding

What about reviews of agency factfinding? Here, empiricism has thus far neglected to tread. Paul Verkuil has noted that “scope of review over facts, including application of law to facts, remains a neglected activity” among legal scholars.¹³⁵ Because no one has empirically evaluated judicial review of agency factfinding in a systematic way over years and agencies, I gathered data from the D.C. Circuit to do so. I identified all of its decisions reviewing agency factual determinations to see whether they were supported by substantial evidence review between 2000 and 2004, dates picked because they were roughly contemporaneous with prior scholarship on judicial review of agency legal determinations and judicial review of agency action for arbitrariness. Once again, substantial evidence review is the standard that applies when agencies engage in formal adjudication; there is a trial-type procedure before an ALJ, followed by review by the agency’s principals.

I obtained from a standard legal database a list of cases in the D.C. Circuit that applied substantial evidence review to a decision by any agency that appeared before it. All told, 30 agencies were sued on this basis, and some of them appeared in court more than others. The NLRB accounted for 99 of the 226 cases, the Federal Energy Regulatory Commission (FERC) was the defendant in 35 cases over this period, and the FCC was involved in 20.¹³⁶

¹³⁵ *Id.* at 682 n.4.

¹³⁶ The D.C. Circuit is not always clear about whether it is, in fact, using the substantial evidence standard to decide its cases, even when it cites to it. In some opinions, it would make reference to boilerplate and mention that the agency decision could not be arbitrary and capricious and had to be supported by substantial evidence. If in these cases, the court’s decision did not turn on a factual determination by the agency, I did not count it as a substantial evidence case, though making those determinations was not always easy (they only comprised a few of the 263 cases considered). I also removed from our dataset cases in which the agency’s action was deemed supported by substantial evidence but illegal for some other reason. Because these cases were not agency wins, it seemed incongruous to treat them as such simply because they won on the substantial evidence review, but lost for some other reason. In other, easier cases, the court would mention the term “substantial evidence” but not base its review on it; I excluded these cases as well. Finally, I checked to see that any case that cited 5 U.S.C. § 706(2)(E), the part of the APA that defines the substantial evidence standard, was included in my database (they all were). Nonetheless, it is worth noting that these judgment calls could have an impact on the results here (although not by very much one way or the other). All told, 38 of the 264 cases initially collected, or 14% of them, were discarded for these reasons.

The bottom line lies in the rate in which the court affirmed the agency. In substantial evidence cases, the agency was affirmed in 71.2% of these cases, a figure slightly higher than, but not dramatically different from, the agency's overall win rates in *Chevron* cases and *State Farm* cases before the appellate courts.

For the purposes of this Article, this basic affirmance rate is the most interesting result in the data. The critical observation is that the affirmance rates under the various standards of judicial review of agency action are quite similar, suggesting that courts may use similar degrees of scrutiny regardless of the doctrinal basis of the review.

We can see that substantial evidence review is not different from the other sorts of review, even though it is concerned with agency factfinding, while the other standards are concerned with agency legal interpretations or their general arbitrariness. The result is consistent with the theme in this Article, which is that courts may be approaching judicial review in a more general fashion than the fact that there are six different standards of review for agency action might suggest. Again, the bottom line is that the real world of judicial review might better be encapsulated by positing that there is a reasonableness standard for agency action that applies regardless of the doctrinal test to be used.

Nonetheless, the prior literature has unpacked the aggregate numbers; I do so here as well to see whether the eerie consistency in validation masks other variables that can more precisely explain how agencies fare in the courts of appeals. That literature has concerned itself in particular with the question of how politicized judicial review is—whether liberal and conservative judges treat agencies differently. Veteran observers of the D.C. Circuit have also suggested that the judges on that court treat some agencies differently from others.¹³⁷ I found little—but not zero—evidence of ideologically polarized voting in substantial evidence review cases, and found that the appellate court does appear to reverse agencies

¹³⁷ See, e.g., Patricia M. Wald, The “New Administrative Law”—With the Same Old Judges In It?, 1991 Duke L.J. 647, 662 (“In my twelve years on the D.C. Circuit, I have reviewed hundreds of administrative actions by the [EPA], the [FERC], [OSHA], the Department of Health and Human Services (HHS), the [FCC], the National Labor Relations Board (NLRB), and other agencies. . . . I hold suspicions or impressions about who is on top or at the bottom of the regulatory honor roll . . .”).

haled before it frequently more than agencies who are rarely sued. This effect, however, was not significant when subjected to logit regression analysis.

To evaluate the data for evidence of politicization, I counted the votes of the individual judges in these cases and assembled a file of case-by-judge observations. Of the 678 total votes cast by judges in these cases, 377 were votes by appointees of Republican presidents, and 301 were votes by appointees of Democratic presidents.

I focused on the party of the appointing president as a measure of ideology because those data are interesting in their own right and because other studies on appellate review of administrative decision making have used them as a proxy for the ideology of the judge, as, for example, in the work most familiar to legal scholars, including that of Miles and Sunstein, Richard Revesz, and Tracey George.¹³⁸ There are other measures for ideology of appellate judges; for example, the judicial common space (JCS) score uses the party of the president and the party of the home state senators as a proxy for ideology.¹³⁹ But JCS scores strike many law professors as odd, and less graspable than the party of the appointing president, and some of the other, finer measures of ideology developed by political scientists—the Martin-Quinn scores or the Maltzman-Bailey citation method, for example—have not yet been applied to the appellate courts during the years of this study.¹⁴⁰

¹³⁸ See Tracey E. George, *Developing a Positive Theory of Decisionmaking on U.S. Courts of Appeals*, 58 *Ohio St. L.J.* 1635, 1640 (1998); Miles & Sunstein, *Chevron*, supra note 123, at 848; Richard L. Revesz, *Environmental Regulation, Ideology, and the D.C. Circuit*, 83 *Va. L. Rev.* 1717, 1727 n.31 (1997) (“The views generally held by the party of the President who appointed a judge were used as a proxy for that judge’s ideology.”).

¹³⁹ Michael W. Giles, Virginia A. Hettinger & Todd Peppers, *Picking Federal Judges: A Note on Policy and Partisan Selection Agendas*, 54 *Pol. Res. Q.* 623, 631 (2001). Lee Epstein and her co-authors have explained how the scores work: “If a judge is appointed from a state where the President and at least one home-state Senator are of the same party, the nominee is assigned the NOMINATE Common Space score of the home-state Senator (or the average of the home-state Senators if both members of the delegation are from the President’s party). If neither home-state Senator is of the President’s party, the nominee receives the NOMINATE Common Space score of the appointing President.” Lee Epstein et al., *The Judicial Common Space*, 23 *J.L. Econ. & Org.* 303, 306 (2007).

¹⁴⁰ The Maltzman-Bailey approach uses citation to place judges on an ideological spectrum, and is, as of this writing, unpublished; Martin-Quinn uses judicial votes, among other factors. See Andrew D. Martin & Kevin M. Quinn, *Dynamic Ideal Point*

In aggregate, the judges were quite similar, regardless of the party of the president who appointed them. Republican-appointed judges voted to uphold the agency determination 70% of the time. Democrat-appointed judges upheld the agency in this period 72% of the time. On the simple question of whether agencies would win the case, their record in substantial evidence disputes is consistent with the way appellate judges evaluated agencies under *Chevron*, in which Democratic-appointed judges upheld the agency as often as did their Republican counterparts.¹⁴¹ In *State Farm* review, the difference reported by Miles and Sunstein was greater (10%), but not outlandishly large.¹⁴²

This does not mean that politics do not play a role in judging agency factfinding. There is no question that they do in many cases, although neither the descriptive statistics nor the regression analysis suggests politics play an important role specifically in substantial evidence review. The implication may be that—at least for fact-based cases—there is a range of acceptable agency behavior. Within that range, agencies may find the facts they like without much ideological policing by judges, either because fact-based cases are viewed as cases for smaller stakes than are law-based cases (where there is slightly more evidence of ideological policing, though not a higher reversal rate), or because ideological commitments match unevenly with factual findings. Judicial scrutiny may vary in predisposition, but the fact is that in litigated substantial

Estimation via Markov Chain Monte Carlo for the U.S. Supreme Court, 1953–1999, 10 *Pol. Analysis* 134, 137 (2002). For a glowing recommendation of the Martin-Quinn scores by a legal scholar, see Ward Farnsworth, *The Use and Limits of Martin-Quinn Scores to Assess Supreme Court Justices, with Special Attention to the Problem of Ideological Drift*, 101 *Nw. U. L. Rev.* 1891, 1892–96 (2007). Epstein et al., have recently attempted to develop JCS scores for all appellate judges; their study, however, ends at the beginning of 2000, the first year of the study undertaken here. Their study essentially *does* apply the Martin-Quinn scores to appellate judges, but, again, outside the time of this study. See Epstein et al., *supra* note 139, at 306–07.

¹⁴¹ Miles & Sunstein, *Chevron*, *supra* note 123, at 849 (“The validation rate of the average Democratic appointee is higher than that of the average Republican appointee by only 0.03 of a percentage point, which is an insignificant difference.”). Miles and Sunstein, however, attempted to evaluate whether a decision was liberal or conservative in these cases, and concluded that it was in those decisions where the politicization of the judiciary was most apparent. *Id.* at 849–50.

¹⁴² Miles & Sunstein, *Arbitrariness*, *supra* note 109, at 776 (“[U]nder arbitrariness review the validation rate of Democratic appointees is 10 percentage points higher than that of Republican appointees.”).

evidence cases, judges may view factfinding as less rooted in policy than are legal interpretations or arbitrariness reviews. Instead, agencies usually win their substantial decisionmaking reviews—but not more often than they do for any determinations they make.

The evidence of politicization lies in an unpacking of the data, though even here, at least on the D.C. Circuit during this period, the evidence is hardly jaw-dropping. Consider the experiences of one agency, often thought to be a politicized one: the NLRB, in which the competing interests of labor unions and interests of corporations are funneled through a closely divided board. The various judges who considered NLRB decisions on substantial evidence review during this period ranged in affirmance rates from 100%, for then-Judge Roberts, to 37.5%, for Judge Silberman. Democratically appointed judges affirmed NLRB substantial evidence findings at a rate of 65%, while Republican appointed judges did so only at a rate of 60%. In NLRB cases where a corporation was the plaintiff (that is, where the agency decision was not pro-business), however, Republican judges voted for the agency only 53% of the time. The question becomes whether these agency-, judge-, and plaintiff-related factors are the real variables of interest in teasing out how judges treat agencies in substantial evidence review.

Because prior efforts to measure judicial voting in administrative law cases attempted to deal with the politicization of adjudication, I also collected data on politicization for substantial evidence cases. But it is worth noting that even if these cases are politicized and ideological voting is a fact of life on the courts of appeals, the purpose of this Article is not to delve into that oft-studied question. Instead, this Article is concerned with doctrine qua doctrine and whether it is really the case that the doctrinal distinctions made by the courts of appeals are holding up.

Nevertheless, it could be the case that liberal judges and conservative judges select very different agency decisions to affirm even though they affirm agencies at roughly similar rates. Given the paucity of dissents on the D.C. Circuit—they appeared in only 8 of the 226 cases in my data—it is unlikely that ideological differences are particularly stark in substantial evidence cases, which is something of a contrast to prior scholarship on affirmances in *Chevron*

cases.¹⁴³ Some prior scholarship has teased out a larger ideological effect by matching liberal and conservative judges with liberal and conservative decisions. Miles and Sunstein, for example, have coded the underlying agency action for its liberalness, and found that therein lies a difference between liberal and conservative judges.¹⁴⁴ Cross and Tiller did the same in their early, influential study,¹⁴⁵ and I do so here as well—albeit cautiously.

Miles and Sunstein admit that their coding method for the agency decision—whether the plaintiff was a corporation or industry group, in which case the underlying agency decision was presumed to be liberal—is imprecise.¹⁴⁶ It is a particularly rough method to apply to the fact-finding data here.¹⁴⁷ In my database, thirty agencies were sued, and the fact that a corporation was the plaintiff does not predict whether the agency action were liberal in each of these areas of administrative law. (For example, consider whether the disbarment of a stockbroker is liberally pro-regulatory or conservatively tough on quasi-crime.) Moreover, in the business regulation context, small businesses may have different interests from large ones, and some non-corporate plaintiffs may be pursuing ends that members of the Federalist Society would find congenial. Among FERC plaintiffs, for example, government-owned utilities are difficult to characterize as pro-business or pro-consumer plaintiffs. Nonetheless, this inexact guide is posited by some to be the research standard.¹⁴⁸ It also serves as some evidence as to whether a court is pro-business or not, as opposed to “conser-

¹⁴³ See Miles & Sunstein, *Chevron*, supra note 123, at 849–50.

¹⁴⁴ See *id.* at 829–31 for a description of their effort to tease out the liberalness of decisions.

¹⁴⁵ Frank B. Cross & Emerson H. Tiller, *Judicial Partisanship and Obedience to Legal Doctrine: Whistleblowing on the Federal Courts of Appeals*, 107 *Yale L.J.* 2155, 2169 (1998).

¹⁴⁶ Miles & Sunstein, *Chevron*, supra note 123, at 831 (“The advantage of this method of coding decisions is its mechanical and objective character. The disadvantage is its crudeness.”).

¹⁴⁷ Miles and Sunstein have another measure of ideology: whether the agency decision were made during a Democratic or Republican presidential administration. Miles & Sunstein, *Arbitrariness*, supra note 109, at 775 (“Our second proxy for the political direction of the agency decision is whether the case was decided during a Democratic or Republican administration.”).

¹⁴⁸ See Frank B. Cross, *Political Science and the New Legal Realism: A Case of Unfortunate Interdisciplinary Ignorance*, 92 *Nw. U. L. Rev.* 251, 290–91 (1997).

vative”—in the sense of holding conservative social values. At any rate, most administrative law cases are brought by businesses: 170 of the 226 substantial evidence cases in my sample featured corporate plaintiffs.

Accordingly, I do not put much stock in the measure. For the sake of scholarship comparability, however, I coded the cases to see whether they were brought by corporate plaintiffs on the assumption that if courts reversed agencies more often in those cases than in others, they might be displaying some anti-regulatory zeal. As it turned out, agencies won 70% of the cases where corporations were not plaintiffs and 72% of the cases where corporations were plaintiffs. The difference was not statistically significant.

Although panel effects—that is, the number of Republican judges on a panel—have mattered in some other contexts, it did not make a difference on the substantial evidence review cases in my database. Agencies won 69% of the cases in which only Democratic appointed judges served on the panel, 72% of the cases where there were two Democratic appointed judges on the panel, 70.2% of the cases where there was only one Democratic appointed judge on the panel, and 73% of the cases in which no Democratic appointed judge served. The differences were, once again, not statistically significant.

Finally, although separate opinions are rare in the D.C. Circuit, it was worth noting the effects on agency success when these sorts of opinions were written. While none of these outcomes were statistically significant, judges were more likely to write concurring opinions when agencies lost (they did so in 7.7% of those cases) than when they won (they did so in 4.3% of those cases). Twelve of the 226 cases featured concurrences. As for dissents, the results were not even suggestive: during the period of study, two dissents were written in cases where the agency lost (representing 3.1% of agency-loss cases); six dissents were written when the agency won (3.7% of agency-win cases).

There are some other interesting aspects of judicial review for substantial evidence worth noting. The purely descriptive data suggested that familiarity breeds contempt. The most common substantial evidence litigant, the NLRB, won 63% of the votes cast during this time period. Similarly, FERC won 66% of the votes. These vote shares were lower than the vote shares received by

other, less frequent litigants. The four agencies that appeared before the court more than ten times (in order from most to least: the NLRB, FERC, the FCC, and OSHRC) won their cases 68% of the time. By contrast, the agencies that appeared before the D.C. Circuit fewer than 10 times during this period were affirmed at an 80% rate.

It is impossible to know whether these results were driven by judicial tastes on the D.C. Circuit, or by selection of cases by litigants, though the results suggest a probable combination of these two variables. Litigants may choose to bring cases against FERC and the NLRB more often because they have learned they are more likely to win those cases, which may in turn be driven by increasing judicial skepticism about the cases. In addition, the more frequently sued agencies decide more matters through formal adjudication than do some other, less frequently sued agencies. And their cases may be more valuable to prospective litigants than are the disputes managed by the biggest player in formal adjudication: the SSA, which employs roughly 900 of the 1300 ALJs working for the federal government.¹⁴⁹ SSA ALJs decide the Administration's cases, particularly disability benefit cases, which have their own limitations on judicial review, and may not be worth taking through the federal courts.¹⁵⁰

As it turned out, however, a closer statistical analysis revealed that none of these factors—*none*—made a statistically significant difference in the way judges decided substantial evidence cases.

¹⁴⁹ For some background on the lot of ALJs and the prominence of the SSA in employing them, see Robin J. Arzt, David H. Coffman & Pamela L. Wood, Federal Administrative Law Judges Conference, *Advancing The Judicial Independence And Efficiency Of The Administrative Judiciary: A Report to the President-Elect of the United States*, 29 *J. Nat'l Ass'n Admin. L. Judiciary* 93, 95 (2009) ("There are over 1,300 ALJs assigned to 31 different agencies, of which the [SSA] is by far the largest employer.").

¹⁵⁰ For a description of this review, which is done under a rather specified substantial evidence standard and performed by the district court, see Paul R. Verkuil & Jeffrey S. Lubbers, *Alternative Approaches to Judicial Review of Social Security Disability Cases*, 55 *Admin. L. Rev.* 731, 738–42 (2003). Before that review commences, they travel through the administrative judiciary in vast numbers. Margaret H. Taylor, *Refugee Roulette in an Administrative Law Context: The Déjà vu of Decisional Disparities in Agency Adjudication*, 28 *J. Nat'l Ass'n Admin. L. Judiciary* 158, 173 (2008) ("SSA receives some five million disability applications each year, and its ALJs resolve almost 500,000 contested cases.").

In order to assess whether the agency sued, the number of Republican judges on the panel, the existence of a separate opinion (concurrence or dissent), or the possible corporate identity of the plaintiff affected the chance of winning a case, I conducted a number of statistical tests.¹⁵¹ First, I examined each of the variables marginally. None of the thirty agencies demonstrated a statistically significant difference in win rate from the overall win rate for all cases. Whether there were zero, one, two, or three Republicans on the panel, the win rate did not differ from the overall win rate. In fact, it was remarkably equal in these four scenarios. Similarly, the win rates for cases with a concurring opinion or dissenting opinion did not differ, nor did they when I examined whether the plaintiff was a corporation.¹⁵² In addition to these marginal tests, I conducted a multivariate test by entering all of the above variables into a logistic regression. None of the variables was statistically significant in this model, nor was the full model itself—that is, a regression including all of these variables.

For substantial evidence, then, none of the ideological, environmental, or other factors affected the overall validation rate, at least not in a statistically significant way, within the given dataset. Although the aggregation strategy pursued here has its limits—sometimes, as appears to be the case in some studies of *Chevron* or judicial review in specific issue-areas, a win rate masks real ideological divisions that would help lawyers better evaluate their prospects than would reference to validation alone¹⁵³—it is nonetheless striking in its own right. Sometimes win rates really do tell you all you need to know about cases, and the baseline prospects for the agency make for an important first step in evaluating any case.

Moreover, the prospects of ideological judging, agency track records, and all the rest of the variables deemed important in others' studies, of course, have almost nothing to do with the precise test

¹⁵¹ These tests were conducted with the assistance of the statistics department at The Wharton School, University of Pennsylvania.

¹⁵² Examination of the four possibilities—no concurring or dissenting opinion, concurring opinion only, dissenting opinion only, or both—was not possible here, since there was only one case with both a concurring and dissenting opinion.

¹⁵³ This is particularly true in the D.C. Circuit, which makes the identity of panels known well in advance of the argument—a practice which has led some savvy D.C. lawyers to abandon claims when confronted with a particularly unfriendly judicial draw.

for judicial review chosen by the appellate court. The politics will presumably play their role no matter what the standard of review or agency sought. Although these other variables do not appear to make a difference here, it is also not clear that they make a difference for the other standards of review because our strategy there is to aggregate across standards—and find remarkable similarity when we do.

CONCLUSION

Observers have always suspected that outcomes under the current system of review are actually determined by factors other than those identified by the courts. In 1941, the Attorney General prepared a report on judicial review in administrative law that posited that many of the relevant considerations for whether an agency won or lost on appeal could only be described as “unarticulated.”¹⁵⁴ The report, submitted to Congress during the realist era of legal scholarship, but still cited today, hints that it may be easy for lawyers and judges to get caught up in the doctrine of judicial review while ignoring other fundamentals.¹⁵⁵

The normative recommendation of this Article is that courts and scholars should focus more on the unarticulated bases for reversal in administrative law and less on the standards of review. While giving the impression of precise analysis, overemphasis of the various standards has actually turned the doctrine of judicial review in administrative law into a morass of complexity. The reality is, in fact, much simpler. Agencies are really reviewed, once the validation rates of the agencies across doctrines taken seriously, under a

¹⁵⁴ The report observed that

In exercising their powers of review, the courts have been influenced, it is commonly thought, by a variety of inarticulate factors: The character of the administrative agency, the nature of the problems with which it deals, the nature and consequences of the administrative action, the confidence which the agency has won, the degree to which the review would interfere with the agency's functions or burden the courts, the nature of the proceedings before the administrative agency, and similar factors.

Report of the Comm. on Admin. Procedure, *Administrative Procedure in Government Agencies*, S. Doc. No. 77-8, at 91 (1941).

¹⁵⁵ See *id.* at 87–92 (expressing concern about the potential detrimental expansion—because of doctrinal ambiguities—of what should be a narrow inquiry into standard of judicial review).

2010]

Reasonable Agencies

187

reasonableness standard that, as the appellate courts currently operate, means that they win cases brought roughly two-thirds of the time.

If this is the case, various intelligent observers, ranging from followers of Edmund Burke to those who think that doctrine evolves efficiently, might wonder why courts have persisted in complicating matters. Before concluding with an explanation of some of the advantages of a simplification of administrative procedure, it is worth exploring the best case for why it has thus far remained so complicated.

A. The Case for Current Doctrine and Its Discontents

The current system of review would not have survived this long without some scholarly and judicial support. Its supposed benefits, however, would not be lost by adopting a single standard of review based on reasonableness.

1. Clarity

One argument for the preservation of the current sextet of standards of review is that they are arguably more “clear” than a reasonableness standard: they define specific guidelines to apply in specific situations. An easily articulated “reasonableness” standard, in contrast, leaves room for ambiguity and, consequently, means less predictability on a case-by-case basis. It is certainly true that a reasonable-agency-action test means, in any given case, something that only fact-specific precedent can tease out. But this only underscores the need for the retention of counsel familiar with what the standard of review means to different judges assessing different agency actions in different factual contexts—something that a well-advised potential litigant certainly would not ignore. Although the interpretive task is not an easy one, it is by no means impossible. Police officers and criminal defense lawyers must assess the reasonableness of conduct in relation to the Fourth Amendment search and seizure thousands of times a year.¹⁵⁶ Negli-

¹⁵⁶ For an overview, see Akhil Reed Amar, *Fourth Amendment First Principles*, 107 *Harv. L. Rev.* 757, 758 (1994); Erin Murphy, *Manufacturing Crime: Process, Pretext, And Criminal Justice*, 97 *Geo. L.J.* 1435, 1470–71 (2009).

gence lawyers make similar assessments of reasonableness in the context of social obligations.¹⁵⁷

2. Judicial Role

A second objection to the adoption of a reasonableness standard might be that it dispossesses judicial review of much of its power, and accordingly does not permit courts to participate fully in the exercise of elucidating and developing our public values.¹⁵⁸ This, of course, is true in any regime of deferential judicial review, and this Article is not the place to visit the voluminous critiques of *Chevron* and substantial evidence for adopting this sort of deference.¹⁵⁹ It is worth noting, however, that substantial evidence is often thought to be an effective standard of review, which I argue is, in the end, nothing more than a reasonableness inquiry, and *Chevron* has been adopted enthusiastically by those judges disempowered by it.¹⁶⁰

3. Agency Discipline

It could be argued that agencies, if they need only to be reasonable interpreters of law, rather than correct interpreters of it, may neglect to pursue their statutory missions to the strictest punctilio. Elizabeth Foote, ordinarily a critic of the ossification that can result from all-too-vigorous judicial review, has suggested that agencies have not in all cases prospered with the flexibility granted to

¹⁵⁷ James Gibson, Doctrinal Feedback and (Un)Reasonable Care, 94 Va. L. Rev. 1641, 1646 (2008). (“And then there is that familiar legal fiction, ‘reasonableness,’ which invites us to use real-world practice as a guide for legal decisionmaking: [t]ort law declares us negligent if we fail to provide ‘reasonable care’ and conform to the conduct of a ‘reasonable person.’”).

¹⁵⁸ The classic exposition of this view is Owen M. Fiss, Against Settlement, 93 Yale L.J. 1073, 1085 (1984).

¹⁵⁹ Indeed, Justice Breyer himself has looked askance at the *Chevron* test on the basis that it deprives judges of the ability to perform the sort of statutory interpretation analyses that they do best. Stephen Breyer, Judicial Review of Questions of Law and Policy, 38 Admin. L. Rev. 363, 377–81 (1986). For a Supreme Court decision suggesting that Breyer has put his doctrinal reservations into practice, see *Christensen v. Harris County*, 529 U.S. 576, 595 n.2 (2000) (Stevens, J., dissenting) (endorsing “fully” Justice Breyer’s narrow reading of *Chevron*).

¹⁶⁰ Indeed, Gary Lawson believes that the Supreme Court had no idea it was fomenting an administrative law revolution when it decided *Chevron*; the case only took hold when the D.C. Circuit leapt to enshrine it as an overarching new standard applicable to most of the rulemaking cases they reviewed. See Lawson, *supra* note 10, at 442–43.

them by *Chevron*'s reasonableness review.¹⁶¹ Foote's concern, too, goes to the age-old debate about administrative flexibility versus administrative constraint, and how one comes down on the debate often turns on what one thinks about public choice theory, regulation in general, and so on. It may, however, be easier for agencies to assess just how much leeway they do have under a simplified standard of review.

4. *The Popularity of Complex Standards of Review*

Underlying all of these concerns is a useful insight: the variety provided by these standards of review is popular and has been embraced, in some quarters, as useful. As Keith Werhan has said, "[t]he APA's indeterminate standards of substantive review reflect Congress's recognition that it is undesirable, and perhaps impossible, to reduce this crucial judicial function to words."¹⁶² Courts are partial to *Chevron*, have cited it extremely often, and appear to appreciate it much more than the even more disjointed standards of review that preceded it.¹⁶³ Fixing substantial evidence review of agency factfinding might be thought of as a solution in need of a problem—that standard, too, has been embraced by both courts and commentators.¹⁶⁴ Many observers are glad that the slippery *State Farm* review of agency policymaking for arbitrariness exists as a backstop protecting the public from extreme agency action, and hypotheticals about agency discretion are a mainstay of most administrative law courses (say, if the EPA insisted that all cars that did not get 80 miles per gallon be taken off the streets, a rule that would both be within its legal authority—it is not permitted to

¹⁶¹ See Elizabeth V. Foote, *Statutory Interpretation or Public Administration: How Chevron Misconceives the Function of Agencies and Why It Matters*, 59 *Admin. L. Rev.* 673, 709 (2007) ("This under-review of administrative work for its reasonableness as bureaucratic action has unfortunate consequences for the quality of public administration.").

¹⁶² Keith Werhan, *Principles of Administrative Law* 310 (2008).

¹⁶³ For a critique of the pre-*Chevron* standards of review, see *United States v. Mead Corp.*, 533 U.S. 218, 240–41 (2001) (Scalia, J., dissenting).

¹⁶⁴ See ABA Judicial Review Guide, *supra* note 91, at 159–60 and accompanying text.

take the costs of a rule into account when assessing its need—and which it could find facts to support).¹⁶⁵

In my view, these inclinations are not attributable to the doctrinal advantages of this or that cabined standard of review, but rather, general approval for the leeway that courts do give agencies—no matter what standard of review they are purporting to apply. But if scholarly and lawyerly approbation for legal tests is based on the fact that, in the end, courts only reverse negligent agencies, then the question of the usefulness of the doctrinal complexity remains—both as a matter of doctrine, and, in light of the empirical evidence, as a matter of what courts actually do. The supporters of various standards of judicial review, the enthusiasts for the *Chevron* test, and the paeans to the usefulness of substantial evidence, probably, at bottom, express approval of the functional outcomes of the complex regime we now have, rather than approval of the complexity itself. Moreover, the most popular tests, as courts have applied them, are those tests—*Chevron* and substantial evidence—that doctrinally do turn on the reasonableness of the agency. In some ways, the approbation for this or that Supreme Court innovation could be read, on the ground, like a guarded suggestion that courts simplify their standards of review. It is precisely this suggestion, this Article posits, that courts have adopted.

B. The Implications of Reasonableness

There are advantages to adopting the reasonableness perspective in administrative law. First among them, of course, is that it actually describes what agencies do. But there are other attractions. The conventional doctrinal story of judicial review in administrative law is distracting. Distinguishing between unreasonable legal interpretations, as is required by *Chevron* step two, and arbitrary administrative decisions, as is required by 5 U.S.C. § 706(2)(B), has never been a satisfying exercise; it is worth coming to terms with the fact that it probably is not a necessary one either.

The doctrinal story that distinguishes among the standards of review encourages lawyers to focus on form, rather than substance. Agencies, we are told, can act through rulemaking or adjudication,

¹⁶⁵ See, e.g., Merrick B. Garland, *Deregulation and Judicial Review*, 98 Harv. L. Rev. 505, 525–26 (1985).

through formal procedures or informal ones.¹⁶⁶ Much turns, it had been suggested, on the choice of action the agency makes—and, therefore, on the consequential standard of judicial review afforded to each category.¹⁶⁷ But once we recognize that courts do not alter their baselines based on agency action, or based on whether they are reviewing factual interpretations, legal interpretations, or policymaking choices, the rather arid exercise of policing the margins of these decisions and actions by agencies becomes much less important—which will be a salve to both scholars and law students required to parse the categorization of the agency action taken.

I suspect, moreover, that administrative lawyers will have recognized that they should be appealing to judges based on the basic reasonableness—or not—of the agency’s actions.¹⁶⁸ To these observers, who, after all, have skin in the game, it will never be satisfactory to argue the technicalities of this or that standard of review in lieu of an appeal to the justice of the case—an alternative that requires scrutiny of the agency’s action on the basics, rather than the technicalities.

The conclusion, then, is that courts ought to acknowledge in their opinions the actual test they apply. They ought to acknowledge that the standard for agency action, regardless of whether it involves factfinding, legal interpretation, or policymaking, is a reasonableness one.

¹⁶⁶ Indeed, their choice of procedures is apparently a matter of absolute discretion. *SEC v. Chenery Corp. (Chenery II)*, 332 U.S. 194, 203 (1947). See Thomas W. Merrill & Kathryn Tongue Watts, *Agency Rules with the Force of Law: The Original Convention*, 116 *Harv. L. Rev.* 467, 492 (2002), for criticism.

¹⁶⁷ *Dickinson v. Zurko*, 527 U.S. 150, 162 (1999) (noting the importance of the various standards of review); Verkuil, *supra* note 5, at 682 (same, albeit noting them somewhat skeptically).

¹⁶⁸ As the experienced appellate advocate Howard Bashman has observed, “lawyers whose appellate briefs fail to explain clearly and convincingly why their client deserves to win on appeal, and why the trial court’s rulings were wrong or right, decrease the client’s chances of winning and complicate the appellate court’s effort to reach the right result.” Howard J. Bashman, *Ten Tips for Excellence in Appellate Advocacy*, *Law.com* (Nov. 13, 2006), <http://www.law.com/jsp/article.jsp?id=1163194123382&hbxlogin=1>. Although I can only source it as hearsay, I have heard my former colleagues in the Department of Justice contend that an important part of their appellate strategy often comes down to a recommendation that the court should “trust us, we’re the government.”

Moreover, for those worried that a simplification of these standards would result in some sort of an abnegation of doctrine, it is worth noting that it is very likely that the reality of judicial review in administrative law is matched in other areas of appellate review. As Richard Posner has observed in the context of civil procedure:

Appellate review is . . . intuitive, though judges pretend otherwise. Opinions recite a variety of standards of review—plenary, clearly erroneous, substantial evidence, some evidence, a modicum of evidence, reasonableness, arbitrary and capricious, abuse of discretion, *Chevron*, *Skidmore*, and so forth—but the gradations of deference that these distinctions mark are finer than judges want, can discern, or need. The only distinction the judicial intellect actually makes is between deferential and nondeferential review. . . .

So what is involved in appellate review is, at bottom, simply confidence or lack thereof in another person's decision. That is an intuitive response informed by experience with similar decisions. It is not rule- or even standard-driven, except in the clearest cases, but it is not mindless guesswork either.¹⁶⁹

Indeed, Posner believes that this reasonableness may be defended as normatively desirable. He thinks that “the judge's proper aim in difficult cases is a *reasonable* result rather than a demonstrably right one.”¹⁷⁰ In administrative law, it is not even clear that there is much difference in analytical rigor between hard looks and other kinds of looks, at least not as the courts are actually practicing review today, as I have discussed earlier. Posner's admonition is, I think, a partial recommendation for simplification.

Amid the morass of complicated standards of review, the inclination of judges to do the reasonable thing will be strong, and—as Posner himself exemplifies—there might be something about using a reasonableness standard in judging that appeals to judges themselves. In fact, the standard might appeal both to the novice and the veteran judges in administrative law.

After all, judges unfamiliar with a particular area of administrative law are likely to resort to first principles in adjudicating over it

¹⁶⁹ Richard A. Posner, *How Judges Think* 113–14 (2008).

¹⁷⁰ Richard A. Posner, *The Problems of Jurisprudence* 26 (1993) (emphasis added).

rather than first mastering the technicalities of, say, common carrier rules governing payment systems in telephony (as is the case with the rather complex system of managing the relationship between local and long distance providers governed by the Telecommunications Act of 1996, and administered by the FCC)¹⁷¹ or rail transport (which has long been the subject of complex ratemaking proceedings by first the Interstate Commerce Commission and now a bureau of the Department of Transportation).¹⁷² It is easier to focus on reasonableness and process than the particulars of the policies in these cases, and that in turn weighs heavily in favor of a reasonableness inquiry, even if it is *sub silentio* a reasonableness inquiry.

Alternatively, veteran hands at administrative law—say, the judges on the D.C. Circuit, may be just as inclined to turn to their horse sense, rather than to every particular of a case, when passing judgment. These judges have experience with almost every agency, and all-but-relationships with the most common litigants, and there too, there is every reason to think that a sense of rough justice may animate decisions in particular cases, rather than a turn to every jot and tittle of the administrative record.

Recognition of this fact, and the turn to an explicit, rather than an implicit, adoption of a reasonableness standard of review would ease the burden of opinion-writing. Judges would not have to twist themselves into knots discerning among various standards of review, as the D.C. Circuit has found itself doing in a variety of cases.¹⁷³ And the Supreme Court would not need to spend much of its limited docket deciding where, exactly, *Chevron* applies and where it does not—a doctrinal area that has been something of a growth industry for the Court ever since the 1982 decision.

Instead, if one relatively graspable standard is the standard that courts really apply, be it to questions of fact, questions of law, or to the arbitrariness of agency policymaking, lawyers and laypersons can focus on the things that do matter to courts, which may make judicial review more predictable and transparent.

¹⁷¹ Pub. L. No. 104-104, sec. 401, § 10, 110 Stat. 56, 128–29 (1996) (codified as amended at 47 U.S.C. § 160 (2006)).

¹⁷² Act to regulate commerce, ch. 104, 24 Stat. 379 (1887) (codified as amended at scattered sections of 49 U.S.C.).

¹⁷³ See *supra* notes 54–56, 103–105 and accompanying text.

Moreover, a reasonableness standard of review can bring to the fore some of the things that agencies do that always exercise judicial attention. For example, reversals on appeal may be particularly likely when the agency reverses the factual findings of its hearing officers, but the substantial evidence standard of review hardly makes this clear; a broader reasonableness standard would provide a basis for understanding why courts so act.¹⁷⁴ Reversals are also thought to be particularly likely where agencies failed to explain the basis of their decisions, especially when those decisions are reversals from prior practice. The need for explanation is something that mattered in both the *Overton Park* and the *State Farm* cases, but a duty to explain has never been easy to analogize to terms like “arbitrariness” on which those doctrines putatively turn.¹⁷⁵ And the importance of explanation divided the Supreme Court in, for example, last term’s *FCC v. Fox Television Stations, Inc.*, in which the adequacy of the agency’s explanation of its decision to ratchet up its policing of curse words for offensive conduct split the justices, an example of the special care that agencies must take when they change their mind.¹⁷⁶

It may also be the case that the scale of the agency action matters. Aggressive *Chevron* interpretations in minor matters may not interest appellate judges or justices, but when the FDA presented a perfectly reasonable interpretation of its authority to regulate drugs and devices that would have permitted it to regulate cigarettes, the Supreme Court rather famously balked.¹⁷⁷ In *Brown & Williamson*, the Court appeared to conclude that even a reasonable interpretation of broad powers to regulate, if it was not vindicated by a clear congressional mandate when novel supervision of a large

¹⁷⁴ This was the case both in *Universal Camera Corp. v. NLRB*, 340 U.S. 474, 491–97 (1951), and in a number of well-known appellate cases that intimate that such reversals occasion a closer look by the appellate judges. See, e.g., *Harberson v. NLRB*, 810 F.2d 977, 984 (10th Cir. 1987) (reversing agency for failing to articulate the reasons for departing from recommendation of ALJ); *Int’l Bhd. of Elec. Workers v. NLRB*, 510 F.2d 1274, 1276 (D.C. Cir. 1975) (same); see generally Michael Asimow, *A Guide to Federal Agency Adjudication* 93–94 (2002) (describing an agency’s ability to reverse initial decisions, even on factual matters).

¹⁷⁵ See supra notes 97–99 and accompanying text.

¹⁷⁶ 129 S. Ct. 1800, 1812–13, 1817 (2009); *id.* at 1824 (Kennedy, J., concurring in part and concurring in the judgment); *id.* at 1826–28 (Stevens, J., dissenting); *id.* at 1833–36, 1838–41 (Breyer, J., dissenting).

¹⁷⁷ *FDA v. Brown & Williamson Tobacco Corp.*, 529 U.S. 120, 125–26 (2000).

industry is at stake, might present different issues.¹⁷⁸ In that case, the Court concluded that broad delegations of authority were not enough, even if the agency interpretation of the delegation might have been, as a matter of textual interpretation, reasonable: “Congress could not have intended to delegate a decision of such economic and political significance to an agency in so cryptic a fashion.”¹⁷⁹ The surmise is that big stakes cases thus may receive particular attention, while lower stakes cases may be more likely to pass muster.¹⁸⁰ While testing this proposition empirically is beyond the scope of this Article, it makes sense that a bottom-line-oriented judiciary sincerely believes that big cases deserve closer scrutiny. And, if so, a reasonableness standard is more likely to recognize this tendency honestly than are any of the six judicial review standards for agency action that currently exist.

But there are other implications to be drawn. One of the reasons for deference to agencies is purportedly based on their expertise. Early administrative law theorist James Landis praised this expertise as a basis for delegation to the administrative state.¹⁸¹ Justice Brandeis, in *West Virginia v. Pennsylvania*, opined that some government responsibilities require skills that judges do not possess; where “[t]he decisions to be made would be of the character which calls for the informed judgment of a board of experts. . . . Clearly this Court could not undertake such determinations.”¹⁸²

The reasonable agency standard that courts actually apply, however, suggests that this expertise rationale—which has animated courts and scholars for years—is not the definitive basis for judicial deference. Courts, after all, apply the reasonableness standard to all individuals, expert or not, in tort law, to the police in evaluating reasonable suspicion pursuant to the Fourth Amendment, and to agencies as well.¹⁸³ It is not, in other words, that the agencies re-

¹⁷⁸ *Id.* at 131–33, 159, 161.

¹⁷⁹ *Id.* at 160.

¹⁸⁰ ABA Judicial Review Guide *supra* note 91, at 179–80.

¹⁸¹ James M. Landis, *The Administrative Process* 154–55 (1938) (praising the professionalization of the bureaucracy, and concluding that it entitled them to flexibility in interpreting their legislative remit).

¹⁸² *Pennsylvania v. West Virginia*, 262 U.S. 553, 621–23 (Brandeis, J., dissenting).

¹⁸³ *Terry v. Ohio*, 392 U.S. 1, 21 (1968) (“The scheme of the Fourth Amendment becomes meaningful only when it is assured that at some point the conduct of those charged with enforcing the laws can be subjected to the more detached, neutral scru-

ceive their generous flexibility before the courts because they are experts, but rather it is because they are presumed to be average, but designated with a particular task to do. And so courts, in practice, give agencies some flexibility in performing the tasks with which they have been charged; the basis for the flexibility appears to lie less in administrative puissance and more in the legal choice by Congress to invest agencies with decisionmaking powers.

Finally, the fact that agencies get reversed, regardless of the standard of review employed, at roughly the same rates, with some allowances made for the political predispositions of the judges, panel effects, and the identity of the plaintiff, suggests that there is something to be said about the rule of law in administrative law. Selection bias—that is, the willingness of potential litigants to bring some cases but not others—may play a role as well, but, as I have explained, the possibility that different selections lead to such similar reversal standards is unlikely (although impossible to dismiss entirely).¹⁸⁴

Legalization may not turn on complex legal doctrines to which courts carefully adhere. But the culture of judicial review may suggest a rough consistency between the ways that Democrat and Republican judges review agency action, regardless of the standard of review applied.

The insight has implications for some of the more creative prescriptions devised to deal with ideological voting on the courts of appeals. While Richard Revesz's observations regarding the pres-

tiny of a judge who must evaluate the reasonableness of a particular search or seizure in light of the particular circumstances.”). There are, of course, reams of articles on this issue, but for considerations more cognate to this Article, see Amar, *supra* note 156, at 801 (“The core of the Fourth Amendment, as we have seen, is neither a warrant nor probable cause, but reasonableness.”); Orin S. Kerr, *The Fourth Amendment in Cyberspace: Can Encryption Create a “Reasonable Expectation of Privacy?”*, 33 *Conn. L. Rev.* 503, 523–24 (2001) (discussing the relationship between reasonableness and probability).

¹⁸⁴ Although one advantage of studying the courts of appeals is that they do not select their cases, at least not as much as does the Supreme Court, whose case law is particularly “fraught with . . . selection effects.” Jonathan Masur, *Judicial Deference and the Credibility of Agency Commitments*, 60 *Vand. L. Rev.* 1021, 1058 n.162 (2007) (concluding that, nevertheless, empirical research based on the Court’s output may be useful); see also Eskridge & Baer, *supra* note 128, at 1096 (“Relatedly, the agency-interpretation cases that come before the Supreme Court are not representative of the cases that come before the courts of appeals.”).

ence of ideological voting on the D.C. Circuit may suggest a reason to get rid of “hard look” arbitrariness review,¹⁸⁵ it is not clear that in the aggregate hard look review is treated differently from any other sort of review—or that judges of different ideologies, at least on the appellate courts in administrative law cases, really approach cases so fundamentally differently. A similar point might be made about Miles and Sunstein’s suggestion that perhaps all panels ought to include ideologically dissimilar members in *State Farm* cases.¹⁸⁶ Courts both increasingly are moving to a reasonableness evaluation in administrative law, and, as I have explained, have many reasons for doing so. Reasonableness is a coin of the realm in many areas of judicial review; the question posited by this Article is whether appellate courts reviewing administrative law decisions should act differently from the many other situations in which they have adopted the reasonableness standard. There is no reason to treat appellate courts’ review of administrative decisions as exalted, complex mechanisms when the true operation of judicial review is clear both in practice and, increasingly, in doctrine. The reasonableness standard in administrative law makes no such mistake.

¹⁸⁵ See Revesz, *supra* note 138, at 1717–19, 1769–71 (1997).

¹⁸⁶ *Dickinson v. Zurko*, 527 U.S. 150, 162 (1999) (noting the importance of the various standards of review); Verkuil, *supra* note 5, at 682 (same, albeit noting them somewhat skeptically).