

DUBIOUS DEFERENCE: REASSESSING APPELLATE STANDARDS OF REVIEW IN IMMIGRATION APPEALS

*Michael Kagan**

ABSTRACT

The long-standing doctrine of deferential review by appellate courts of findings of fact by administrative agencies is seriously flawed for two main reasons. First, the most prominent justification for deference relies on the empirical assumption that first-instance adjudicators are best able to determine the truth because they can directly view witness demeanor. Decades of social science research has proven this assumption about the value of demeanor false. Second, in principle, the deference rule applies to all types of administrative adjudication, with no attention to the relative gravity of interests at stake in different types of cases or to the varying levels of actual expertise that different executive agencies bring to bear. These weaknesses are particularly acute in immigration appeals and help explain why the 2002 streamlining of the Board of Immigration Appeals has proven problematic for the federal courts. Appellate courts often take advantage of the inherent ambiguities of the deference doctrine to prevent unacceptable results, but this approach does little to repair the essential flaws in the doctrine and exposes courts to criticism that they are acting arbitrarily. A more coherent way to understand how appellate courts use deference in practice would be to apply a balancing analysis similar to the procedural due process doctrine.

TABLE OF CONTENTS

INTRODUCTION	102
I. OVERVIEW OF THE DEFERENCE DOCTRINE	107
A. <i>De Novo v. Deference</i>	107
B. <i>Origins of Deference in Administrative Law</i>	113
C. <i>Justifications for Deference</i>	114
D. <i>Systemic Doubts</i>	118
II. THE ACCURACY RATIONALE	125
A. <i>The Demeanor Assumption</i>	127

* Associate Professor, University of Nevada Las Vegas, William S. Boyd School of Law. I am grateful for the input of Linda Berger, Bill Ong Hing, and Katherine Kruse, and for the research of Matthew Wright.

B. <i>The Comparative Role of Juries, Judges, and Agencies</i>	137
III. THE EFFICIENCY RATIONALE	139
A. <i>An Independent Rationale for Deference?</i>	139
B. <i>Streamlining and Deterrence</i>	141
IV. DELEGATION OF POWERS.....	145
V. THE INARTICULATE FACTORS	150
A. <i>Agency Expertise</i>	150
B. <i>The Interest at Stake</i>	155
VI. DEFERENCE AS A BALANCE.....	158

INTRODUCTION

In 2002, Attorney General John Ashcroft enacted significant changes in the way the Board of Immigration Appeals (BIA) decides immigration cases, precipitating a dramatic increase in the number of immigration cases reaching the federal courts of appeal.¹ These appeals, which within five years constituted 18% of the total federal appellate docket nationally and around a third of the workload of the Second and Ninth Circuits,² were followed shortly by expressions of skepticism from some federal judges about the quality of immigration administrative adjudication.³ Some courts have issued searing attacks on the basic competence of immigration adjudicators while expressing concern about resulting tension between the administrative agencies and the courts.⁴ The Seventh Circuit's Chief Judge Posner has been particularly energetic in these criticisms, to the extent that a series of Seventh Circuit decisions have been interpreted by some as an attempt to carve out an exceptional doctrine of appellate review specific to immigration cases.⁵ The post-2002 prob-

1. See John R. B. Palmer et al., *Why Are So Many People Challenging Board of Immigration Appeals Decisions in Federal Court? An Empirical Analysis of the Recent Surge in Petitions for Review*, 20 *Geo. Immigr. L.J.* 1, 3 (2005) (noting that the courts of appeal received about five times more immigration-related petitions after the 2002 BIA changes).

2. Adam Cox, *Deference, Delegation and Immigration Law*, 74 *U. CHI. L. REV.* 1671, 1683–84 nn. 41–42 (2007).

3. Eric M. Fink, *Liars and Terrorists and Judges, Oh My: Moral Panic and the Symbolic Politics of Appellate Review in Asylum Cases*, 83 *NOTRE DAME L. REV.* 2019, 2020–21 (2008).

4. See, e.g., *Benslimane v. Gonzales*, 430 F.3d 828, 829–30 (7th Cir. 2005) (“This tension between judicial and administrative adjudicators . . . is due to the fact that the adjudication of these cases at the administrative level has fallen below the minimum standards of legal justice. Whether this is due to resource constraints or to other circumstances beyond the Board’s and the Immigration Court’s control, we do not know, though we note that the problem is not of recent origin.”). For additional examples, see *infra* Part IV.

5. See, e.g., Cox, *supra* note 2, at 1684.

lems with immigration appeals have motivated thoughtful proposals for structural reform of immigration adjudication⁶ and raised urgent empirical questions about why more immigrants are choosing to challenge their deportations in the federal courts.⁷

Yet, someone deeply immersed in fundamental principles of American litigation might be surprised that the 2002 BIA reforms generated any trouble. The 2002 reforms of the BIA and several Congressional actions targeting immigration appeals since then were based on two central and long-standing pillars of American administration of justice. First, Ashcroft's changes were built around the distinction between questions of law and questions of fact. Second, Ashcroft bolstered the rule that first-instance findings of fact should be reviewed with deference on appeal.⁸ The Attorney General prescribed that the BIA should reverse the factual findings of immigration judges only if clearly erroneous.⁹ Ashcroft also eliminated the requirement that the BIA write a persuasive opinion in each case¹⁰ and allowed most appeals to be decided by single-member panels.¹¹ The principle that first-instance decisions should be treated with deference was thus given concrete procedural form.

The Attorney General justified the 2002 expansion of deference in BIA appeals by referencing the standards "now commonly used by the federal courts" and, for justification, cited the Federal Rules of Civil Procedure.¹² The Attorney General further looked to the Supreme Court's decision in *Anderson v. City of Bessemer*,¹³ which was heard under the rules of civil procedure and not by an administra-

6. See Stephen H. Legomsky, *Restructuring Immigration Adjudication*, 59 DUKE L.J. 1635 (2010); Fatma E. Marouf, *Implicit Bias and Immigration Courts*, 45 NEW ENG. L. REV. 417 (2011).

7. See, e.g., Palmer et al, *supra* note 1, at 82-85.

8. See 8 U.S.C. § 1252(a)(2) (2006) (eliminating judicial review in removal of certain aliens who have committed certain criminal offenses, but providing "[n]othing in subparagraph (B) or (C), or in any other provision of this chapter (other than this section) . . . shall be construed as precluding review of constitutional claims or questions of law raised upon a petition for review.").

9. 8 C.F.R. § 1003.1(d)(3)(i).

10. Legomsky, *supra* note 6, at 1657.

11. *Id.*

12. See Board of Immigration Appeals: Procedural Reforms to Improve Case Management, 67 Fed. Reg. 54878, 54889-90 (Aug. 26, 2002) [hereinafter BIA Reforms] ("Just as the Supreme Court has concluded that on balance the 'clearly erroneous' standard is an effective, reasonable, and efficient standard of appellate review of factual determinations by federal district courts, the Department has concluded that the 'clearly erroneous' standard is an effective, reasonable, and efficient standard for appellate administrative review of factual determinations by immigration judges.") (citations omitted).

13. 470 U.S. 564, 574-75 (1985); see also BIA Reforms, *supra* note 12.

tive agency of any kind. The fact that the Attorney General referenced civil procedure in writing rules for administrative adjudication should perhaps have raised more questions. But given that deference to findings of fact, and the fact-law dichotomy itself, are so widely accepted in our legal system, it is perhaps not surprising that the Attorney General imported them to the specific arena of immigration adjudication with little analysis.

Deference to first-instance findings of fact is so deeply rooted in the American legal tradition that it is easily assumed as an unquestioned part of the legal landscape. According to well-established law, findings of fact by administrative agencies should be treated with deference on appeal. Courts rarely question this rule, and in the last ten years both the Executive Branch and Congress acted to bolster its application in the field of immigration adjudication. And yet, though the deference doctrine has deep roots in our legal tradition, the reality that strict application of this rule has caused problems should come as no surprise.

Appellate deference in adjudication of individual administrative cases¹⁴ has generally received less attention than deference to administrative rule-making,¹⁵ commonly known as *Chevron* deference.¹⁶ My purpose in this Article is to look more closely at this doctrine. The deference doctrine asks appellate judges to affirm factual findings by executive agencies even when the judges believe the agency is likely to be wrong, and even when the human costs of allowing a factual error to stand would be extreme. Appellate deference to factual adjudication by administrative agencies should be considered disturbing on its face and should not stand without a compelling justification. This jarring rule emerged from a specific political context, fed by ancient mythologies about human capaci-

14. The Administrative Procedure Act ("APA") defines an "order and adjudication" as follows: "'Order' means the whole or any part of the final disposition (whether affirmative, negative, injunctive, or declaratory in form) of any agency in any matter other than rule-making but including licensing. 'Adjudication' means agency process for the formulation of an order." Administrative Procedure Act of 1946, Pub. L. No. 79-404, § 2(d), 60 Stat. 237, 237 (1946); see also Kenneth Culp Davis, *An Approach to Problems of Evidence in the Administrative Process*, 55 HARV. L. REV. 364, 402 (1942) ("When an agency finds facts concerning immediate parties—what the parties did, what the circumstances were, what the background conditions were—the agency is performing an adjudicative function, and the facts may conveniently be called adjudicative facts.").

15. The APA defines a rule as "the whole or any part of any agency statement of general or particular applicability and future effect designed to implement, interpret, or prescribe law or policy or to describe the organization, procedure, or practice requirements of any agency." § 2(c), 60 Stat. at 237.

16. *Chevron, U.S.A., Inc. v. Nat'l Res. Def. Council, Inc.*, 467 U.S. 837, 843 (1984).

ties, and has been shielded from scrutiny by judges willing to make grandiose assumptions about human nature simply because other judges have made such assumptions before. These mythologies and assumptions have been thoroughly undermined by decades of empirical research by social scientists, and the deference doctrine is thus overdue for reconsideration.

At the same time, significant doubts have been raised about whether stated rules about deferential standards of review accurately describe what courts of appeal actually do in administrative appeals. A series of quantitative studies have shown the various standards of review used in administrative appeals in federal court have no clear impact on the results of the cases. This aggregate quantitative finding encourages the perception that while appellate decisions routinely state the standard of review, and judges and parties often focus considerable energy on it, “what courts are really doing is the same sort of analysis regardless of the standard of review.”¹⁷

A separate line of scholarship puts these empirical doubts in context, showing that judicial flexibility was implicitly built into the deference doctrine from its earliest days. The deferential standard of review on questions of fact developed from particular pressures on the Supreme Court at a particular historical moment before World War I. To escape difficulties raised by turn-of-the-century railroad regulation cases, the Supreme Court transposed to administrative law the fact-law distinction that originally had its roots in the unique role of juries in our system of justice. The Court did this at a time when American legal culture favored formalistic and categorical rules, even as the Court was searching for flexibility. The resulting deference doctrine is thus at once broad in reach, superficially inflexible, and yet ambiguous in terms of practical application. The same substantial evidence standard of review applies in fact-finding in everything from asylum to union organizing. But its boundaries—determining what is law and what is fact—and its real meaning—determining what constitutes substantial evidence and what constitutes clear error—are constantly unclear, thus facilitating what Justice Frankfurter called “[s]ome scope for judicial discretion.”¹⁸

There may yet be good reason to apply deference in some cases, but it is difficult to justify the across-the-board requirements for deference that continue to be part of our black-letter law. My argument

17. See David Zaring, *Reasonable Agencies*, 96 VA. L. REV. 135, 169 (2010).

18. See *Universal Camera Corp. v. NLRB*, 340 U.S. 474, 489 (1951).

is that the key to understanding why deference is not always easy to apply is to focus on whether it actually can be justified in a specific type of adjudication. Sometimes there is reason to believe that an administrative adjudicator has an advantage in reaching an accurate understanding of the facts but often there is not. Some administrative agencies have real expertise that courts learn to trust, while other agencies give courts reason to doubt their decisions. Sometimes there may be a reason to seek judicial efficiency even at the cost of accuracy, but sometimes not. While the deference doctrine on its face does not openly embrace these factors, courts often do take them into account in how they apply the doctrine. Specifically in the context of immigration cases, especially those immigration cases where the stakes are high, the usual justifications for deference are not persuasive.

When justifications for deference are made more specific to particular types of cases, and when other factors like the weight of the interest at stake are taken into account, a balancing test similar to the doctrine of procedural due process, as developed by the Supreme Court in *Mathews v. Eldridge*, emerges.¹⁹ I will argue that it is quite logical that this form of analysis has applicability to appellate standards of review, and that it would helpfully discipline the flexibility practiced by judges. Balancing tests provide a structured way for courts to adapt to different circumstances without relying on a hidden hand of unexplained discretion. Weighing a balance of factors is what many courts already often do when they decide how much deference to give an agency's finding of fact, and for good reason. I will make the case that standard of review doctrine would be rendered less ambiguous, more analytically predictable, and more in tune with empirical knowledge about how the world works if courts incorporated a *Mathews*-like balancing test more directly. Applying this analysis in the immigration context illustrates why deference is a poor fit for this type of case.

Part I provides an overview of the deference doctrine, the justifications usually given for it, and the challenges that it has faced in application. Part II examines the role of what I will call the "demeanor assumption" in supporting deferential review. Part III examines the premise that a deferential appellate standard of review promotes judicial efficiency. Part IV examines the delegation of the powers doctrine, which in administrative law can be a powerful rationale for judicial restraint. In Part V, I examine two "inarticulate

19. 424 U.S. 319 (1976).

factors” often mentioned by courts as a rationale for grading the level of deference up or down—namely, the relative expertise of different executive agencies—and the weight of the interest at stake in a particular case. Part VI explains why it is sensible to reframe the deference doctrine as a balancing test that would more directly incorporate these factors and justifications, while allowing courts to adapt both to different types of cases and to evolving understandings about the empirical validity of long-standing assumptions.

I. OVERVIEW OF THE DEFERENCE DOCTRINE

A. *De Novo v. Deference*

Justice Stephen Breyer recently explained the increasing importance of judicial review of agency decision-making:

Whether we like it or not, government administration is everywhere. . . . Our legal system asks courts to review agency work because the technical nature of modern society, along with the public’s desire for Social Security, medical care, and the like, has brought laws that delegate enormous decision-making power and responsibility to administrators who are not themselves elected. . . . How can we ensure that related administrative decisions are fair and reasonable? Or, as the ancient Romans put it, *quis custodiet ipsos custodes?* Who will regulate the regulators?²⁰

This basic challenge of modern democracy is the seed of modern administrative law. And at the center of administrative law is the idea that courts should often defer to executive agencies even as they review the decisions the agencies make. The most oft-discussed form of this is *Chevron* deference, which applies to agency rule-making and interpretation of statutes. The focus here is on review of agencies’ adjudication of facts in individual cases. But to look closely at judicial review of factual adjudication by agencies, it is important to look beyond the field of administrative law. For one thing, the rules of deference in administrative law are an application of a more general idea in our legal system: appellate courts should adjust their standard of review not only in different cases, but for different questions within the same case. The most classic of these distinct standards of review is the separation of questions of law

20. Stephen Breyer, *The Executive Branch, Administrative Action, and Comparative Expertise*, 32 CARDOZO L. REV. 2189, 2190–92 (2011).

from questions of fact. In administrative law, there are typically six main formulations of the standard of review, some intended for statutory and rule-making matters and others for factual issues.²¹ But there are up to thirty different formulations of judicial deference in the American legal system more generally.²²

I will provide a brief summary of three standards that are particularly important for this article—*de novo*, substantial evidence, and clearly erroneous—before highlighting the central concept of deference. The simplest and most important standard of review as a point of reference is *de novo* review. Under *de novo* review, an appellate judge will affirm if she agrees with the decision below and will substitute her own judgment if not.²³ *De novo* review is the easiest to understand because the result on appeal should correspond precisely to what the appellate judge considers the most correct decision. In this article, I use *de novo* review as a baseline. By contrast, other standards of review ask appellate judges to uphold decisions notwithstanding doubts about their correctness. I include any such standards under the category deferential review.

As we shall see, courts imported the basic idea of deference to agency findings of fact from civil and criminal procedure, where appellate courts apply a deferential standard of review to the first-instance trier of fact. In administrative law, this was translated to the substantial evidence standard or reasonableness review, which asks if there was sufficient evidence to allow a “reasonable mind” to support the conclusion.²⁴ This has been explained as a showing that less than a preponderance of the evidence, but “more than a mere scintilla,”²⁵ supports the decision. As one court elaborated:

[T]he question whether an agency determination is supported by substantial evidence is the same as the question whether a reasonable fact finder could make such a determination based upon the administrative record. If a reasonable fact finder could make a particular finding on the administrative record, then the finding is supported by sub-

21. See Richard J. Pierce, Jr., *What Do the Studies of Judicial Review of Agency Actions Mean?*, 63 ADMIN L. REV. 77, 78–83 (2011).

22. Richard H. W. Maloy, “Standards of Review” – *Just a Tip of the Icicle*, 77 U. DET. MERCY L. REV. 603, 610 (2000); see also Jonathan Armiger, Note, *Judicial Review of Public Utility Commissions*, 86 IND. L.J. 1163, 1168 (2011).

23. CHARLES H. KOCH, JR., 3 ADMINISTRATIVE LAW AND PRACTICE, 90–94 (3d ed. 2010).

24. *Universal Camera Corp. v. NLRB*, 340 U.S. 474, 477 (1951).

25. *Consol. Edison Co. of N.Y. v. NLRB*, 305 U.S. 197, 229 (1938); KOCH, *supra* note 23, at 335–36.

stantial evidence. Conversely, if no reasonable fact finder could make that finding on the administrative record, the finding is not supported by substantial evidence.²⁶

As a standard for reviewing findings of fact, the substantial evidence standard is easily confused with the clearly erroneous standard,²⁷ under which an appellate court must have a “definite and firm conviction” to remand or reverse because there is “a fairly strong presumption in favor of the inferior authority.”²⁸ The substantial evidence standard is more common to administrative law²⁹ because it is used in the Administrative Procedure Act (APA),³⁰ while clearly erroneous is more tied to civil cases, enshrined as it is in the Federal Rules of Civil Procedure.³¹ But despite the difference in terminology, there has been significant blurring of the distinction, raising questions about whether there is any real substantial difference between the two standards. Substantial evidence is sometimes described as being “more deferential” than the clearly erroneous standard,³² a distinction drawn originally from the difference between reviewing a jury verdict as opposed to factual findings by a judge.³³ Yet, even courts that try to draw a distinction between these formulas minimize it, noting on the one hand that any witness credibility decision is entitled to special deference, while on the other hand suggesting that substantial evidence review in administrative cases may be *less* deferential than in jury cases, even with regard to credibility.³⁴ In the final analysis, it is possible clearly erroneous and substantial evidence may be two sides of the same coin. On its face, one standard is positive: a decision should be upheld if supported by substantial evidence. The other standard is negative: a decision should be reversed if it is clearly an error. These are not

26. *Dia v. Ashcroft*, 353 F.3d 228, 249 (3d Cir. 2003).

27. FED. R. CIV. P. 52(a)(6); KOCH, *supra* note 23, at 331.

28. KOCH, *supra* note 23, at 331.

29. *Id.* at 336–37.

30. 5 U.S.C. § 556(d) (2006).

31. FED. R. CIV. P. 52(a)(6).

32. *Chen v. Mukasey*, 510 F.3d 797, 801 (8th Cir. 2007); *see also* KOCH, *supra* note 23, at 331 (describing clearly erroneous review as a weak form of de novo review).

33. *Chen*, 510 F.3d at 801; *see also* Paul R. Verkuil, *An Outcomes Analysis of Scope of Review Standards*, 44 WM. & MARY L. REV. 679, 693–95 (2002) (analyzing how the Supreme Court considers the scope of review standards).

34. *See, e.g., Chen*, 510 F.3d at 801 (“Because of these principles of administrative law, ‘substantial evidence’ review of administrative findings entails review of an IJ’s credibility determinations, whereas ‘substantial evidence’ review of a jury’s findings defers almost entirely to the jury’s credibility determinations.”) (internal citations omitted).

contradictory descriptions. It is difficult to imagine a decision that one could reasonably say is a clear error but yet is somehow substantially supported by evidence.

Adding to this confusion, the Executive Branch has imported the clearly erroneous standard into administrative immigration appeals, even though substantial evidence is the more common standard in administrative cases.³⁵ The Board of Immigration Appeals reviews credibility findings by immigration judges through the clearly erroneous standard.³⁶ The Attorney General justified this rule by referencing the standards “now commonly used by the federal courts” and cited the Federal Rules of Civil Procedure as well as the Supreme Court’s decision in *Anderson v. City of Bessemer*, which I will discuss in greater depth in Part III.³⁷ *Anderson* was an employment discrimination case heard under the Federal Rules of Civil Procedure with findings of fact by a federal district court, not an administrative agency.³⁸

The application of two differently-phrased deferential standards of review at the BIA and then at the federal court of appeal can produce somewhat peculiar results. At the most mundane level, a lawyer filing an appeal of a finding of fact to the administrative Board of Immigration Appeals should reference the clearly erroneous standard, but, if the same case is later appealed to the court of appeals, the briefs should be edited to reference the substantial evidence standard instead.³⁹ More substantively, one might think that this double layer of deference would make appeals harder to win at

35. 8 C.F.R. § 1003.1(d)(3)(i) (2012) (prescribing that the Board of Immigration Appeals should apply the clearly erroneous standard of review to factual decisions by immigration judges).

36. *Id.*

37. BIA Reforms, *supra* note 12. (“Just as the Supreme Court has concluded that on balance the ‘clearly erroneous’ standard is an effective, reasonable, and efficient standard of appellate review of factual determinations by federal district courts, *see Anderson*, 470 U.S. at 574–75, and FED. R. CIV. P. 52(a), the Department has concluded that the ‘clearly erroneous’ standard is an effective, reasonable, and efficient standard for appellate administrative review of factual determinations by immigration judges.”).

38. The practical result of this is that a lawyer filing an appeal of a finding of fact to the administrative Board of Immigration Appeals should reference the clearly erroneous standard, but, if the same case is appealed to the court of appeals, the substantial evidence standard should apply instead. *See INS v. Elias-Zacarias*, 502 U.S. 478, 481 n.1, 484 (1992) (using the substantial evidence standard in federal court review of factual questions in an asylum case). But it is debatable whether this change connotes a meaningfully different standard, or if it is analogous to the way substantively similar procedural motions may go under different names in state courts as opposed to the federal courts.

39. *Elias-Zacarias*, 502 U.S. at 481 (using the substantial evidence standard in federal court review of factual questions in an asylum case).

the higher levels since the Court of Appeals is supposed to be deferential to a decision that is itself deferential to another decision. But it turns out that this double layer of deference may actually produce *less* deference. The Tenth Circuit Court of Appeals has understood this scenario to pose the question of whether the BIA properly applied the clearly erroneous standard.⁴⁰ This is a question of law, and thus leads the Court of Appeals to apply a *de novo* standard of review.⁴¹ At least two other circuits have agreed.⁴² In other words, a federal court need not be deferential in deciding whether an administrative tribunal was properly deferential. The Second Circuit has allowed the BIA to escape this trap by re-labeling what might be thought of as a factual matter as a legal question instead, so that no deference would apply.⁴³

The clearly erroneous/substantial evidence comparison is just one example of a wider confusion about the distinction between the different standards of review that might be called on by an appeals court reviewing a finding of fact. There is considerable reason to doubt whether so many gradations of deference can meaningfully be differentiated from each other in practice. Consider, for instance, the difference between substantial evidence on the one hand, and arbitrary and capricious or abuse of discretion review on the other. The APA prescribes that a reviewing court should set aside agency decisions made without a formal hearing if the decisions are “arbitrary, capricious, an abuse of discretion, or otherwise not in accordance with law.”⁴⁴ In theory, substantial evidence and clearly erroneous are less deferential than abuse of discretion and arbitrary and capricious standards.⁴⁵ But in practice, courts do not always specifically explain why they adopt one set of words to describe the standard of review rather than another, and the distinctions between

40. *Kabba v. Mukasey*, 530 F.3d 1239, 1244 (10th Cir. 2008).

41. *Id.* at 1245.

42. See *Kaplun v. Attorney General*, 602 F.3d 260 (3d Cir. 2010); *Alvarado de Rodriguez v. Holder*, 585 F.3d 227, 233 (5th Cir. 2009).

43. See *Xiu Qin Huang v. Holder*, 455 Fed. App'x. 67, 72 (2d Cir. 2012) (“The BIA has not reviewed *de novo* any of the IJs’ factual findings. Instead, the BIA has concluded, on *de novo* review, that the factual findings do not meet the legal standard of an objectively reasonable fear of persecution. . . .”); cf. *Kaplun*, 602 F.3d at 269 (“[I]nsofar as the BIA interpreted 8 C.F.R. § 1003.1(d)(3) to hold that an IJ’s assessment of the probability of future torture is not a finding of fact because the events have not yet occurred, we conclude its interpretation plainly errs.”).

44. Administrative Procedure Act of 1946, Pub. L. No. 79-404, § 10(e), 60 Stat. 234, 243 (1946).

45. *Armiger*, *supra* note 22, at 1169.

them can be easily lost.

As an example, consider a line of labor law cases from the Sixth Circuit. In 1972, the Sixth Circuit said that a factual finding by the National Labor Relations Board (NLRB) could be upheld if backed by “sufficient evidence” so as to not be an abuse of discretion,⁴⁶ and then in 1989, simplified this to “may only be reversed if the Board abused its discretion.”⁴⁷ Then the Sixth Circuit added the proviso that a NLRB finding would be upheld if it is not an abuse of discretion and “the Board’s findings are reasonable.”⁴⁸ This then morphed into an explicit use of the substantial evidence standard.⁴⁹ Only at this last stage did the Sixth Circuit make reference to the fact that there was, and in fact had been all along, a controlling Supreme Court precedent applying the substantial evidence standard in a NLRB case.⁵⁰ But it would be a significant stretch to claim that the circuit court was defying the Supreme Court in any meaningful way for all these years. While the Sixth Circuit was somewhat loose in its terminology, it was hardly trying to change the substance. Rather than imposing different standards of review in different cases, the court was explaining the same basic concept in different ways. The real error may be the tendency to turn such variations of phraseology into distinct doctrinal prescriptions about how appellate courts should do their work.

Commentators have suggested that even at the Supreme Court level the different prescriptions for levels of standard of review in different types of cases have not operated as binding precedents, but are better understood as jurisprudential canons expressing general principles and commitments.⁵¹ It may be a distraction to seek any better parsing of the various formulations and better to focus on the core difference between deferential review (no matter how it is

46. *Michigan Hosp. Serv. Corp. v. NLRB*, 472 F.2d 293, 296 (6th Cir. 1972) (“From the record before us we find sufficient evidence to support the findings of the Regional Director, as adopted by the Board, and we are unable to rule that the Board abused its discretion in designating the unit as appropriate for collective bargaining.”).

47. *Twin City Hosp. Corp. v. NLRB*, 889 F.2d 1557, 1560 (6th Cir. 1989); *see also* *NLRB v. Sweetwater Hosp. Ass’n*, 604 F.2d 454, 457 (6th Cir. 1979).

48. *Randall v. NLRB*, 638 F.2d 957, 959 (6th Cir. 1981).

49. *DTR Indus. Inc. v. NLRB*, 39 F.3d 106, 110 (6th Cir. 1994); *see also* *NLRB v. V & S Schuler Eng’g, Inc.*, 309 F.3d 362, 367 (6th Cir. 2002).

50. *Universal Camera Corp. v. NLRB*, 340 U.S. 474, 491 (1951), *cited in* *DTR Indus., Inc.*, 39 F.3d at 110.

51. Connor N. Raso & William N. Eskridge, Jr., *Chevron as Canon, Not a Precedent: An Empirical Study of What Motivates Justices in Agency Deference Cases*, 110 COLUM. L. REV. 1727, 1727 (2010).

named) and de novo review.⁵² The central point, as stressed by the Sixth Circuit in one of its NLRB cases, was that the “[B]oard’s reasonable inferences may not be displaced on review even though the court might justifiably have reached a different conclusion, had the matter been before it de novo.”⁵³ This is the essential difference between deferential review and de novo review,⁵⁴ and, as long as this distinction is clear, we may leave for another day the question of whether different types of deferential review can be distinguished from each other.

B. Origins of Deference in Administrative Law

The close connection between the civil and administrative standards of review is no accident, as Thomas Merrill has recently shown in a historical study. In the nineteenth century, judicial review of administrative findings of fact was generally de novo, with circuit courts directly receiving evidence and developing their own record, primarily to overrule the Interstate Commerce Commission (ICC) in cases concerning railroad rates.⁵⁵ When a popular backlash calling for stronger regulation of the railroads led to the Hepburn Act in 1906, the Court abruptly began deferring to the ICC’s findings.⁵⁶ In *Illinois Central Rail Road Company v. ICC*, the Court explained its new restrained approach by reference to the distinction between questions of fact and questions of law in common law cases.⁵⁷ Thus, the fact-law dichotomy became the seed for an entirely new relationship between the federal courts and administrative agencies in which appeals in administrative cases are understood as if analogous to

52. Justice Breyer has acknowledged skepticism about whether the different standards of review really can have distinct meanings, but argues, “[j]udges are able to apply different standards – at least to some degree.” Breyer, *supra* note 20, at 2194; *see also* Dickinson v. Zurko, 527 U.S. 150, 153–64 (1999) (finding the substantial evidence standard, not the clearly erroneous standard, applies in patent and trademark cases). For a commentary of the administrative law jurisprudence of Justice Breyer and Justice Scalia, *see* Verkuil, *supra* note 33, at 693–96.

53. NLRB v. St. Francis Healthcare Ctr., 212 F.3d 945, 952 (6th Cir. 2000) (quoting V & S ProGALV v. NLRB, 168 F.3d 270, 275 (6th Cir. 1999)); *cf.* Maloy, *supra* note 22, at 611 (“[With de novo review] [t]he reviewing court’s objective is not limited to ascertaining whether the lower court erred; it intends that the proper party will prevail; hence, it has sometimes been described as a *right/wrong* standard of review.”).

54. KOCH, *supra* note 23, at 329.

55. Thomas W. Merrill, *Article III, Agency Adjudication, and the Origins of the Appellate Review Model of Administrative Law*, 111 COLUM. L. REV. 939, 950–51 (2011).

56. *Id.* at 960.

57. 206 U.S. 441, 466 (1907); *see also* Merrill, *supra* note 55, at 960–61.

appeals in civil or criminal cases.⁵⁸

During this period, American jurisprudence was still heavily rooted in traditions of formalism and categorical reasoning.⁵⁹ This helps explain why, in seeking a new approach toward administrative agencies, the Court fell back on a categorical distinction and created a single, formal rule ostensibly to govern all forms of agency adjudication. These early administrative law battles about the degree to which judges should defer to executive agencies foreshadowed New Deal-era struggles about government intervention in the economy.⁶⁰ But by the late 1930s, critiques of legal formalism had gained ground over more categorical approaches.⁶¹ However, although the Court turned toward more flexible balancing approaches to new legal questions,⁶² the more formalistic deference doctrine remained. I will return to the implications of this shift from categorical reasoning to balancing tests in the final part of this article.

C. *Justifications for Deference*

Deferential review and de novo review will produce identical appellate results in many cases. It is only in one type of case that deference should impact the result. That is when an appellate court believes the decision below is likely an error, but not clearly an error. It is this category of case, and only this category, that makes deferential review more than a mere theory. The following chart illustrates the results that will result depending on how strongly an appellate court considers that a lower decision is incorrect, showing the key difference between de novo and deferential review.

58. Merrill, *supra* note 55, at 960–61.

59. *See id.*; T. Alexander Aleinikoff, *Constitutional Law in the Age of Balancing*, 96 YALE L.J. 943, 949–52 (1987).

60. *See generally* Aleinikoff, *supra* note 59, at 952–58.

61. *See id.* at 955–59.

62. *See id.* at 958–71.

2012]

DUBIOUS DEFERENCE

115

Appellate Court Belief About the Correctness of a Decision:	CORRECT	MAYBE/LIKELY INCORRECT	DEFINITELY INCORRECT
DE NOVO REVIEW	Affirm	Reverse/remand	Reverse/remand
DEFERENTIAL REVIEW	Affirm	<u>Affirm</u>	Reverse/remand

The practical impact of deference can be that the central, high-stakes issues in an immigration case barely receive any analysis on appeal. Consider the impact in a case recently handled by the Thomas & Mack Legal Clinic at the University of Nevada Las Vegas. A man said that he feared deportation in part because of a medical condition that an expert had testified would increase his risk of being tortured in detention if he returned to his country. The condition had been confirmed in one medical report but was not mentioned in a report from another physician. The immigration judge concluded that the man did not really have the condition—a finding that became one of the central issues on appeal to the BIA. But the Board, reviewing under the deferential clearly erroneous standard, dismissed the entire issue, stating, “[t]he Immigration Judge gave specific and persuasive reasons why she discounted this testimony. On review, we do not find any clear error.” The central factual question of whether a man would be in danger of torture was thus dispatched in two sentences.

Consider also a 2011 asylum case in the First Circuit in which Marius Stanciu said that he had been twice detained and beaten by police in his native Romania—on one occasion leading him to be hospitalized for several days—because of his Romani (Roma) ethnicity.⁶³ He asked for asylum for fear of similar violence if he returned home. The immigration judge (IJ) found that Romanis are subject to violence and discrimination in Romania and that Mr.

63. *Stanciu v. Holder*, 659 F.3d 203, 204–06 (1st Cir. 2011).

Stanciu's testimony was "plausible."⁶⁴ But the IJ concluded his claim that he had been attacked so frequently and severely was not credible due to three inconsistencies in his testimony.⁶⁵ The First Circuit thought the credibility assessment was "a judgment call that a reasonable fact finder could resolve either way."⁶⁶ In the end, the court simply concluded that the IJ's finding "has some basis in the record,"⁶⁷ and rested on the rule of deference to "[t]he judge who heard the testimony firsthand [who] always has an advantage."⁶⁸ The court never said what its conclusion about Mr. Stanciu's credibility would have been if it were deciding the case *de novo*.

This First Circuit decision was entirely correct in terms of doctrine, and the IJ may have been correct in terms of the facts of the case. But it raises a critical question that is too rarely asked. Federal courts of appeals judges are among the most esteemed of all judicial personnel in the United States. A panel of three such judges heard an appeal from a man who might have been in danger of being detained and beaten because of his ethnicity, where his fate depended on making a correct judgment about his veracity. In the end, these judges never said what they actually thought about the central issue in his case, except that it was a close call. Instead, they deferred to a much lower-level administrative adjudicator. Their decision leaves open the possibility that they might have been more willing to believe Mr. Stanciu than the IJ had been, or that the panel may have split. We will never know because current doctrine does not consider these judges' actual opinions on the substantive factual finding necessary to resolve an appeal like this. It is this unique category of cases that makes deferential review special and which calls out for justification. Why should we ever demand that a superior court leave in place a decision by an inferior court that it may believe to be wrong?⁶⁹

64. *Id.* at 205–06.

65. He said he had been hospitalized, while his wife said he had only been given medication and recovered at home; he later said he had been disoriented after the beating. He was contradictory about his travel history in and out of Romania. And he described a period of detention once as ten days long, but later said it was six or seven days long. *Id.* at 206–07.

66. *Id.* at 208.

67. *Id.*

68. *Id.* at 206.

69. For an even more vivid example, consider a 2001 Federal Court of Australia case, involving an asylum claim where a man said he would be subject to execution if returned to Iran. *W148/00A v. Minister for Immigration & Multicultural Affairs* [2001] 185 ALR 703 (Austl.). The administrative tribunal had rejected his application because of doubts about his credibility but acknowledged that if he was telling the truth he would face "grave conse-

Courts have offered several alternative answers to this question. The most prominent substantive justification for the wisdom of deference is accuracy, the premise that a first-instance decision maker is better positioned to reach a correct assessment of the facts.⁷⁰ This rationale grows from ancient assumptions in the common law tradition about the value of an adjudicator being able to view a witness live so as to be able to draw conclusions from his or her manner of speaking. By incorporating the fact-law distinction into administrative appeals, the Supreme Court imported into the administrative realm all of the common law's assumptions about how triers of fact make judgments about evidence. This is why, though this Article focuses on administrative cases, the arguments presented here may apply to civil cases where a judge acts as the trier of fact and, to a more limited extent, to jury trials.⁷¹ In the administrative law context, the accuracy rationale may also find support in the idea that executive agencies are experts and thus more likely than generalist judges to reach a correct decision in a case involving technical complexity.⁷² An additional justification was emphasized by the Supreme Court in *Anderson v. City of Bessemer City*⁷³ that deference promotes judicial efficiency by preventing appellate courts from duplicating the work of trial judges.⁷⁴

While the accuracy and efficiency rationales are borrowed from civil law, a third justification applies uniquely to administrative cases. In the late nineteenth century, even before the railroad cases, the Supreme Court had made use of the fact-law dichotomy in its early forays into immigration law. But, in these cases, the Court placed considerable emphasis on Congress's authority to delegate fact find-

quences" upon deportation. *Id.* at ¶ 11. On appeal, the federal court criticized the administrative tribunal's reasoning for being "largely speculative," "inconclusive," and "cause for concern." *Id.* at ¶¶ 67-68. Yet, due to the rule requiring deference to first instance findings of fact, the appellate court affirmed the denial of asylum and allowed the man to be deported. See Michael Kagan, *Is Truth in the Eye of the Beholder? Objective Credibility Assessment in Refugee Status Determination*, 17 GEO. IMMIGR. L.J. 367, 405-06 (2003).

70. *Anderson v. City of Bessemer City*, 470 U.S. 564, 574 (1985).

71. See *infra* note 170 and accompanying text.

72. See *Cincinnati, N.O. & T.P. Ry. Co. v. Interstate Commerce Comm'n*, 162 U.S. 184, 194 (1896) ("[I]t is not necessary for us, in the present case, to express any opinion on a subject so full of difficulty.").

73. 470 U.S. at 574-75.

74. *Id.* at 575; see also *Concrete Pipe & Prods. v. Constr. Laborers Pension Trust*, 508 U.S. 602, 623 (1993) ("[I]n many circumstances the costs of providing for duplicative proceedings are thought to outweigh the benefits . . ."); *Amadeo v. Zant*, 486 U.S. 214, 228-29 (1988) (concerning criminal law); *Inwood Labs., Inc. v. Ives Labs., Inc.*, 456 U.S. 855 (1982) (concerning trademark law).

ing to executive agencies, which then justified judicial restraint.⁷⁵ The delegation of powers doctrine remains a primary justification for deferential review,⁷⁶ and I will explore it in greater detail in Part IV. But it should be understood that delegation of powers is a different kind of justification for deference than the accuracy or efficiency rationales. While congressional delegation can conceivably justify restraint by courts, it does not justify the idea of deference itself. Delegation simply means that this is a choice that Congress can make, but it does not say anything about whether it is a wise choice.

D. Systemic Doubts

Over the last decade, Congress and the Executive Branch have sought to reinforce the principle of deference in immigration adjudication. As we have already seen, the Attorney General's 2002 reforms of the BIA established a clearly deferential standard of review of questions of fact in the administrative appeal stage. In 2005, the REAL ID Act purported to establish a uniform standard for review of witness credibility in immigration cases based on appellate deference.⁷⁷ At the time, Congress expressed concern that some circuit courts were being particularly aggressive about reversing BIA decisions, a concern that has been undermined by empirical analysis.⁷⁸ But statistical evidence does show many circuit courts becoming

75. *Nishimura Ekiu v. United States*, 142 U.S. 651, 660 (1892) (“[C]ongress may, if it sees fit . . . authorize the courts to investigate and ascertain the facts on which the right to land depends. But, on the other hand, the final determination of those facts may be intrusted [sic] by congress to executive officers; and in such a case, as in all others, in which a statute gives a discretionary power to an officer, to be exercised by him upon his own opinion of certain facts, he is made the sole and exclusive judge of the existence of those facts, and no other tribunal, unless expressly authorized by law to do so, is at liberty to re-examine or controvert the sufficiency of the evidence on which he acted.” (citations omitted)). While *Ekiu* expressed this rule in ironclad terms, more recent cases have explained *Ekiu*'s holding as an expression of the usual distinction between questions of law and questions of fact. See *INS v. St. Cyr*, 533 U.S. 289, 312–13 (2001).

76. KOCH, *supra* note 23, at 301 (noting that Congress may delegate adjudication to an administrative agency); see also COX, *supra* note 2, at 1674–77 (suggesting judicial skepticism directed at immigration courts may reflect a “nondelegation norm.”).

77. But the REAL ID Act's approach only lists factors for immigration judges to consider, many of which were already well-established in the jurisprudence before 2005. See e.g., Scott Rempell, *Credibility Assessments and the REAL ID Act's Amendments to Immigration Law*, 44 TEX. INT'L L.J. 185, 206 (2008). The Act does not set out how these factors should be weighed. See *Shrestha v. Holder*, 590 F.3d 1034, 1040 (9th Cir. 2010).

78. Fink, *supra* note 3, at 2035–36, fig. 3 (noting the perception that the Ninth Circuit reversed credibility-based decisions at a higher rate not borne out of statistical data for 1995–2005).

increasingly likely to rule against the BIA after the 2002 BIA changes.⁷⁹ This fact, coupled with the post-2002 explosion in the immigration dockets of the circuit courts, indicates that something about the way deference was applied in immigration cases had unintended results. At the same time, as a matter of black-letter law, the federal courts of appeal continue to review factual findings by the BIA with deference.

The deference doctrine is at once both ambiguous and rigid. It is ambiguous because the exact meaning of the various deferential standards of review is difficult to define with precision, as are the boundaries of the issues to which deference should apply.⁸⁰ As one judge put it, “[t]he concept of deference is admittedly fuzzy, embracing everything from a perfunctory nod to craven acquiescence.”⁸¹ While the basic rules of deferential standards of review have been a constant for many decades, there has long been skepticism about whether consistent application should be expected from courts.⁸² At the same time, the deference doctrine is rigid in the sense that the same basic standard—review of findings of fact for substantial evidence—is applicable across all areas of administrative adjudication and its close cousins are similarly applicable in civil and even criminal cases.

Deference is to a great extent a one-size-fits-all rule, which might prove impractical across such diverse types of adjudication, except that the inherent ambiguities of the doctrine build in enough practical flexibility to keep the rigidity from becoming unworkable. Judges who see a need to adapt to differing circumstances can take advantage of the ambiguity of the rule to avoid an unacceptable result. Much will depend on the judge and his or her motivation to bend the rules to produce a desired result. This trend appears to follow the advice of Justice Frankfurter to implicitly accept a high degree of inconsistency and to trust judges not to manipulate the ambiguity too much:

A formula for judicial review of administrative action may afford grounds for certitude but cannot assure certainty of

79. *Id.* Circuits other than the Ninth began to reverse credibility decisions at a higher rate. “In each of the last two years under examination, the Ninth Circuit had a lower credibility reversal rate than any court other than the Fourth, Fifth, and Eleventh Circuits, none of which ever reversed an adverse credibility determination over the entire eleven-year period from 1995–2005.” *Id.* at 2036.

80. See Maloy, *supra* note 22, at 605–06.

81. FRANK M. COFFIN, ON APPEAL: COURTS, LAWYERING, AND JUDGING 260 (1994).

82. See Verkuil, *supra* note 33, at 685.

application. Some scope for judicial discretion in applying the formula can be avoided only by falsifying the actual process of judging or by using the formula as an instrument of futile casuistry. It cannot be too often repeated that judges are not automata. The ultimate reliance for the fair operation of any standard is a judiciary of high competence and character and the constant play of an informed professional critique upon its work.⁸³

The problem is that not everyone is equally comfortable with judicial discretion as a means of achieving legal flexibility because it places so much weight on the subjective judgment of individual judges.⁸⁴ Certainly, fair application of any legal doctrine requires competent judges of high character, and, whenever a close question arises, the result may depend on the composition of the court. But the deference doctrine poses a different kind of problem because it seems to deliberately put the judicial process in a black box; it claims that there is a universal standard to be applied, while it invites and may even require judges to apply it inconsistently.

There is empirical reason to think that judicial discretion inside the judicial black box extends so far that the black-letter rules mandating different standards of review in different cases no longer have any observable impact on decision-making. A decade ago, Paul Verkuil conducted a quantitative study to try to capture the effects of the varying standards of appellate review on the results of actual cases. The goal was to test the hypothesis that if a standard is more deferential it should produce relatively fewer reversals and remands than a *de novo* standard.⁸⁵ But examining Social Security disability and veterans' disability cases, he found little correlation. In Social Security, around 50% of appeals to federal court—where the sub-

83. *Universal Camera Corp. v. NLRB*, 340 U.S. 474, 488–89 (1951); see also Edward H. Cooper, *Civil Rule 52(a): Rationing and Rationalizing the Resources of Appellate Review*, 63 NOTRE DAME L. REV. 645, 645 (1988) (“This success is due to the fact that the ‘clearly erroneous’ phrase has no intrinsic meaning. It is elastic, capacious, malleable, and above all variable. Because it means nothing, it can mean anything and everything that it ought to mean. It cannot be defined, unless the definition might enumerate a nearly infinite number of shadings along the spectrum of working review standards.”).

84. See, e.g., Thomas J. Miles & Cass R. Sunstein, *The New Legal Realism*, 75 U. CHI. L. REV. 831, 846–49 (2008) (noting that, in *Chevron* deference cases involving environmental and labor law, judges invalidated regulations at a rate of 36%, considered high relative to the expected level of deference, and judges' decisions correlated with whether they were Democratic or Republican appointees and whether the regulation at issue was considered liberal or conservative).

85. Verkuil, *supra* note 33, at 688–91.

stantial evidence standard applied—ended in reversal or remand.⁸⁶ But in veterans' disability, only around a fifth of the appeals to federal court resulted in reversal or remand, even with a standard of review arguably more generous to claimants than the one used in Social Security disability.⁸⁷ These numerical measures do not on their own mean the standards of review are being applied incorrectly. It could be that in the Social Security context people with stronger appeals were more likely to seek review for some reason.⁸⁸ It could also be that there are systemic problems in Social Security disability adjudication.⁸⁹

Yet, over the last two decades, a number of comprehensive empirical studies have been conducted on the impact of different standards of review on the results of appellate decisions. The consistent result is that, if standards of review make any difference, it is difficult to see it statistically. Regardless of what the standard of review may be in various types of administrative law appeals to the federal courts, agencies win on appeal roughly two-thirds of the time (i.e., courts of appeal reverse or remand agency decisions around one-third of the time).⁹⁰ Of particular relevance to this article, a recent study found that when courts of appeal use substantial evidence review, which is applicable to administrative findings of fact, they rule against the agency at roughly the same rate as with other standards of review.⁹¹ Not only is there little or no difference in the observable impact of different formulations of deference, there is little quantitative difference with *de novo* review, as well.⁹² This is a particularly striking finding because, even if the different formulations of deference are impossible to parse, one might still expect deference to be distinguishable from *de novo*. Chief Judge Posner has argued for a two-tiered system of standards of review because "there are more verbal formulas for the scope of appellate review . . . than there are distinctions actually capable of being drawn in the practice of appellate review,"⁹³ but the *de novo* versus deferential

86. *Id.* at 705.

87. *Id.* at 711 n.145. Veterans' law includes a provision for veterans to receive the "benefit of the doubt," though it also encompasses the clearly erroneous standard. *Id.* at 710-11.

88. *Id.* at 706 (noting the case-selection hypothesis).

89. *Id.* at 706-08.

90. Pierce, *supra* note 21, at 84-85; Zaring, *supra* note 17, at 170-76.

91. Zaring, *supra* note 17 at 141, 169, 177-78.

92. Pierce, *supra* note 21, at 83, 85.

93. *United States v. Boyd*, 55 F.3d 239, 242 (7th Cir. 1995); *see also* Armiger, *supra* note 22, at 1179-82 (proposing a three-tiered standard for judicial review of public utility commissions).

distinction “at least is a feasible, intelligible, and important one.”⁹⁴ Yet, even this modest hypothesis has yet to find statistical support.

Correlation (or lack of correlation) does not prove causation. Like Verkuil, the authors of more recent empirical studies note that the lack of statistically observable impact of different standards of review does not rule out the possibility there is still some hidden impact.⁹⁵ The most obvious alternative explanation would be selection bias, whereby would-be appellants are theoretically less likely to file appeals that they are less likely to win due to the standard of review.⁹⁶ Deterring appeals is a potential justification for deferential review, in that it promotes judicial efficiency. This could, in theory, produce a situation where, of those appeals against agency decisions actually filed, the likelihood of success is the same as in *de novo* cases since the weaker appeals are deterred.⁹⁷ Certainly, in individual cases, it appears that judges care a great deal about the standard of review, and it may be that aggregate statistics of standard of review cases may somehow obscure real impacts of deference in different types of cases since judges may in practice apply more real deference in some cases than in others. However, as of now, these alternative hypotheses are plausible but also unproven empirically.

While it would be imprudent to declare definitively that deferential review has no impact, the available empirical evidence raises serious questions about how much impact it really has. The idea of deference suggests an empirically measurable hypothesis that, all other things being equal, courts applying deference should rule in favor of agencies more often than courts that are not applying deference. Since we cannot conduct litigation as a controlled experiment that isolates one single factor, empirical studies may not be able to conclusively disprove this hypothesis. But it is important to note that scholars, to date, have been notably unsuccessful in their search for empirical evidence that the standard of review impacts the results of litigation. As a recent review of the research summarized, “[t]here is no empirical support for the widespread belief that choice of doctrine plays a major role in judicial review of agency

94. *Boyd*, 55 F.3d at 242.

95. *Pierce*, *supra* note 21, at 86-90.

96. Jonathan Masur, *Judicial Deference and the Credibility of Agency Commitments*, 60 VAND. L. REV. 1021, 1058 n.161, 1059 n.166 (2007).

97. It may prove possible to test this theory by examining whether parties are statistically less likely to appeal decisions where they would face a deferential standard of review, but I have not found any studies providing this analysis.

actions."⁹⁸ As of now, with consistent empirical evidence pointing in a single direction, the persuasive burden may be on those who believe standards of review are important to show how they matter in real cases.

When evidence begins to mount that there is a large gap between what judges say and their actual decisions, a doctrine is in trouble. The mildest form of this trouble is the complaint that the doctrinal rules are simply becoming a distraction from the real matter at hand.⁹⁹ David Zaring argues that, since courts appear to apply different standards of review in a similar manner and since standard of review doctrine tends to simply add legal complexity to the appellate process, it would be better to simplify all appellate review of agency decisions to a unitary standard of reasonableness.¹⁰⁰ This form of diagnosis and proposed cure is, in essence, comfortable with judicial discretion, and aims to simplify matters by making the ambiguity of the doctrine more transparent. A single standard of reasonableness—that is no easier to define but at least saves lawyers the trouble of parsing a multitude of indistinguishable deference rules—would replace our current gradations of appellate review, impossible as they are to define with precision. The doctrine-results gap exposed by statistical studies also invites a deeper and more damning critique, one that cannot be addressed by consolidating ambiguity into a more straightforward formulation. What if judges are simply manipulative in the way they use the standard of review doctrine and, for various reasons, simply refrain from expressing their real motivations? Consider again Mr. Stanciu's asylum application in which the First Circuit cited deference in affirming a credibility-based denial without expressing any substantive conclusion about the credibility issue. A cynical reader and, perhaps, a street-smart lawyer, might say that it is clear what the appellate judges thought: they agreed with the immigration judge that Mr. Stanciu was probably a liar, but they just did not say so directly. Had they disagreed with the IJ, they would have said the decision lacked substantial evidence and remanded. But since they agreed, they simply cited deference and moved on. A cynic might say this is why the results of deferential review cases and de novo review cases are the same.

If this cynical explanation is correct even some of the time, there

98. Pierce, *supra* note 21, at 93.

99. See Zaring, *supra* note 17, at 186.

100. *Id.* at 190-97.

are important downsides. First, the deference doctrine deflects the moral burden of making a difficult decision in a high-stakes case from the court of appeals to an immigration judge. The court affirms the decision without saying directly, "We agree." This should be troublesome if one assumes that the honor of being a federal appellate judge carries with it the responsibility to make weighty decisions. Second, by not resolving the central issue in the case, the court restrains itself from offering useful guidance to future immigration judges about how to make difficult decisions in close cases. Deference can have an important systemic downside in that it hinders the gradual case-by-case refinement of law through appellate jurisprudence. The historical fact-law distinction was justified as a way of focusing the energies of appellate courts on cases where their judgments would be most useful as precedents. Deference impedes appellate judges from wrestling directly with the central issue in a case, especially when they agree with the decision below. When a panel splits on a case involving deference, the majority and dissent will not necessarily engage each other's arguments directly since judges voting to affirm are not actually required to say why a decision is right.¹⁰¹ The more general result is that, over time, courts produce a body of decisions listing what would be wrong to do in a credibility assessment, but giving little clear guidance for immigration judges about how to do a credibility assessment correctly.¹⁰² If appellate courts are analogous to teachers, then deference leads them to be the kind of teachers who harshly lecture students about their most severe mistakes, while never showing how to do things correctly.¹⁰³

101. See, e.g., *Toufighi v. Mukasey*, 538 F.3d 988 (9th Cir. 2008). But see *Cole v. Holder*, 659 F.3d 762 (9th Cir. 2011) (engaging in a detailed argument about the evidence where the majority and dissent disagree, *inter alia*, about how much to defer to administrative judgment on how to weigh evidence in a case not involving witness credibility).

102. Cf. Legomsky, *supra* note 6, at 1664 (critiquing the 2002 streamlining of the BIA for encouraging the unexplained affirmation of first-instance decisions, eliminating safeguards that required a decision maker to consider counter-arguments and confirm that a tentative initial impression is actually the correct judgment under law). In particular, the new reliance on single-member panels degrades the long-term development of the law since multi-member appellate panels "permit dissenting opinions that can help steer future law" and facilitate "the exchange of ideas and the airing of differences of opinion." *Id.*

103. For an expansion of this argument in the context of asylum cases, see Kagan, *supra* note 69, at 409 and Steve Norman, *Assessing the Credibility of Refugee Applicants: A Judicial Perspective*, 19 INT'L J. REFUGEE L. 273, 291-92 (2007).

II. THE ACCURACY RATIONALE

The most frequently cited justification for deferring to a first-instance finder of fact is the premise that he or she is in a better position to make an accurate evaluation of the evidence. I call this the accuracy rationale, and much hangs on its validity. If an appellate court can have real confidence that it is deferring to a decision-maker who is most likely to be right anyway, other justifications for deference become secondary. But if the accuracy rationale is shaky—and I will suggest that it is—then alternative rationales must bear much greater weight and greater scrutiny. Any other benefit of deference would then come at the expense of accuracy. If we cannot assume that a first-instance decision is the one most likely to be factually accurate, then it would require an especially compelling reason to defer to it nevertheless.

The accuracy rationale is itself supported by underlying assumptions. One of the oldest in administrative law is the thought that specialized expert agencies are better able to sort through technical complexity.¹⁰⁴ This logic may survive scrutiny better than others, but it is also less far-reaching since, as we will see, not all cases are technically complex and not all executive agencies display the kind of expertise that would support this form of deference. I will explore these issues in detail below in Part V.

The more prominent and far-reaching basis for the accuracy rationale is the one suggested in the Federal Rules of Civil Procedure: “Findings of fact, whether based on oral or other evidence, must not be set aside unless clearly erroneous, and the reviewing court must give due regard to the trial court’s opportunity to judge the witnesses’ credibility.”¹⁰⁵ The central idea here is that the first-instance adjudicator has an advantage in making an accurate decision by observing a witness’s demeanor as he or she testifies in person, and that this advantage cannot be transmitted to an appellate court through a written transcript. I will use the term demeanor to refer to everything about the way a person speaks, except for the actual words that he or she says. I would include here gestures, facial expressions, and tone of voice, among other things. In other words,

104. See *Cincinnati, N.O. & T.P. Ry. Co. v. Interstate Commerce Comm’n*, 162 U.S. 184, 194 (1986) (“Some observations made by the commission, in its report, on the nature of the circumstances and conditions which would justify a greater charge for the shorter distance, gave occasion for an interesting discussion by the respective counsel. But it is not necessary for us, in the present case, to express any opinion on a subject so full of difficulty.”).

105. FED. R. CIV. P. 52(a)(6).

demeanor, as used here, includes all factors that might be relevant to interpreting a witness's testimony that cannot be easily captured on the printed pages of a transcript.¹⁰⁶

There may be reasons other than demeanor to expect more accuracy from first-instance decision makers in some cases. Hearing a live presentation of witnesses may be helpful for a decision maker to digest a particularly complicated or confusing evidentiary record. Among other things, in an oral hearing, he or she enjoys the opportunity to clarify ambiguities resulting from only reading a written transcript or an affidavit.¹⁰⁷ It is for the same basic reason that universities require their students to attend lectures and seminars, rather than limiting their education to assigned reading. Yet, this rationale does not have the broad reach that the demeanor assumption pretends. It gains strength as an evidentiary record grows in complexity and may not apply equally to more simple cases, or even to complex cases where the factual dispute on appeal is fairly narrow.

The idea that people might be able to separate truth from falsehoods by observing demeanor is deeply attractive, with potential benefits in court rooms, police work, business and employment relations, personal relationships, and literally any human interaction where something can be gained or lost by one person convincing another to rely on a particular piece of information. In social science, the pursuit of a technique by which to detect lies through demeanor has been traced to Charles Darwin.¹⁰⁸ In the twentieth century some psychologists grew confident that they could develop a successful method. Sigmund Freud wrote in 1905: "No mortal can keep a secret. If his lips are silent, he chatters with his finger-tips; betrayal oozes out of him at every pore."¹⁰⁹ The idea that social science has already succeeded in developing such techniques is promoted frequently in popular culture,¹¹⁰ including through a recent television

106. One might immediately raise a question about the potential use of audio and video tape. As technological advances make it increasingly more feasible to preserve not just what people say but also how they say it, even demeanor might be reviewable by an appellate court.

107. Mark Spottswood, *Live Hearings and Paper Trials*, 38 FLA. ST. U. L. REV. 827, 850-51, 869 (2011).

108. See Paul Ekman & Maureen O'Sullivan, *From Flawed Self-Assessment to Blatant Whoppers: The Utility of Voluntary and Involuntary Behavior in Detecting Deception*, 24 BEHAV. SCI. & L. 673, 675-76 (2006) (discussing Darwin's 1872 work *THE EXPRESSION OF THE EMOTIONS IN MAN AND ANIMALS*).

109. Charles F. Bond, Jr. & Bella M. DePaulo, *Accuracy of Deception Judgments*, 10 PERSONALITY & SOC. PSYCHOL. REV. 214, 215 (2006).

110. See generally GREGORY HARTLEY & MARYANN KARINCH, *HOW TO SPOT A LIAR: WHY*

drama called *Lie to Me* based on a real researcher, Paul Ekman, whose work I will discuss below.¹¹¹ Nevertheless, social science research into the value of demeanor to lie detection is quite extensive, and the findings not especially encouraging, popular perception notwithstanding.

A. *The Demeanor Assumption*

While social scientists have searched for a way to detect deception through observation of demeanor, our legal tradition has long assumed that people have the ability automatically. I use the term assumption deliberately because it is difficult to find any court that cites any factual basis for the assertion, even though courts make the assertion frequently. The assumption that direct observation of a witness is important surfaced first in the law of evidence to explain the requirement that witnesses should appear at trial in person. It produces peculiar issues in some cases, for instance, on the question of whether a person with impaired eye sight can serve on a jury,¹¹² whether a woman who wants to cover her face for religious reasons should be allowed to testify in court,¹¹³ and whether a witness can testify while wearing sunglasses.¹¹⁴ Perhaps the most prominent promoter of the demeanor assumption, though hardly the earliest, was John Henry Wigmore who, in the course of his studies of evidence, approvingly quoted the Indiana Supreme Court's assertion that "no one who cannot see the expression of faces, nor observe deportment and demeanor, can justly weigh testimony."¹¹⁵

PEOPLE DON'T TELL THE TRUTH . . . AND HOW YOU CAN CATCH THEM (Jodi Brandon ed., 2005); MARK MCCLISH, 10 EASY WAYS TO SPOT A LIAR (2011); BARRY L. MCMANUS, LIAR: THE ART OF DETECTING DECEPTION AND ELICITING RESPONSES (2008).

111. Ekman writes a blog commenting on the degree to which the show, centering around the fictional main character of Dr. Lightman, reflects actual science. The blog includes the caveat: "How the Lightman Group spots lies is largely based on findings from my research. Because it is a drama not a documentary, Dr. Lightman is not as tentative about interpreting behavior as I am. Lies are uncovered more quickly and with more certainty than it happens in reality. But most of what you see is based on scientific evidence." PAUL EKMAN, *Truth About "Lie to Me,"* DR. PAUL EKMAN: CUTTING EDGE BEHAVIORAL SCIENCE FOR REAL WORLD APPLICATIONS, <http://www.paulekman.com> (last visited Oct. 4, 2012).

112. Aaron J. Williams, *The Veiled Truth: Can the Credibility of Testimony Given by a Niqab-Wearing Witness Be Judged Without the Assistance of Facial Expressions?*, 85 U. DET. MERCY L. REV. 273, 275-77 (2008).

113. *Id.* at 273-74 (dismissing a suit because the Muslim female claimant refused to reveal her face during testimony).

114. *Morales v. Artuz*, 281 F.3d 55, 60-62 (2d Cir. 2002).

115. *Rhodes v. State*, 27 N.E. 866, 868 (Ind. 1891).

A leading treatise traces the demeanor assumption to a seventeenth-century English judgment that stated, “[o]ur law requires persons to appear and give their testimony ‘viva voce;’ and we see that their testimony appears credible or not by their very countenances and the manner of their delivery.”¹¹⁶ This idea was repeated by American courts in the nineteenth century. In 1836, the Supreme Judicial Court of Massachusetts wrote: “The false witness cannot endure the stings of his wounded conscience, his countenance and his deportment will, in spite of his endeavors to the contrary, by signs as clear and intelligible as they are inexpressible, declare, that the story which he has just sworn to, is a lie.”¹¹⁷ Other nineteenth-century courts made similar statements, usually without citation, to the effect that a written transcript could be no substitute for live testimony and that people have a natural ability to use demeanor (or “deportment”) to discern the truth.¹¹⁸ An 1857 judgment from the Supreme Court of Missouri referred to the assumed capacity of humans to assess credibility of a live witness as “a vast moral power.”¹¹⁹ Taking the somewhat unusual step of referencing a source for the idea that facial expressions can communicate more than words, the Missouri court quoted from the Gospel of Luke.¹²⁰

With *Mattox v. United States* in 1895, the U.S. Supreme Court began relying on the demeanor assumption to ground its Confronta-

116. JOHN HENRY WIGMORE, 3 A TREATISE ON THE ANGLO-AMERICAN SYSTEM OF EVIDENCE IN TRIALS AT COMMON LAW § 1395 (James H. Chadbourn, rev. ed. 1974) (1923), [hereinafter ANGLO-AMERICAN SYSTEM OF EVIDENCE] (citing Fenwick’s Trial, 13 HOW. ST. TR. 591, 638, 712 (1696)).

117. *Commonwealth v. Richards*, 18 Pick. 434, 439 (Mass. 1836).

118. *See, e.g., Callanan v. Shaw*, 24 Iowa 441, 447 (1868) (“[W]e will believe the one whose appearance, deportment and manner, impress us most favorably. In this, we are guided by our knowledge of human nature, which, in some, appears to be almost intuitive.”); *Sligh v. People*, 11 N.W. 782, 783 (Mich. 1882) (“The production of witnesses in open court is one of the best means of trying their credit, and everyone knows how difficult it is to judge from written testimony of the demeanor and appearance which would strike those who examined them.”).

119. *State v. McO’Blenis*, 24 Mo. 402, 421 (1857) (“Every one must know there is a difference in the effect of the same words when delivered in open court from the mouth of the witness and when read from a deposition. They may seem worthy of credit in the one case, and positively unworthy of all credit in the other. Who can be unmindful of the influence of the manner and carriage of a witness on the stand? There are many things, aside from the literal import of the words uttered by the witness while testifying, on which the value of his evidence depends. These it is impossible to transfer to paper. Taken in the aggregate, they constitute a vast moral power in eliciting the truth, all of which is lost when the examination is had out of court, and the mere words of the witness are reproduced in the form of a deposition.”).

120. *Id.* at 421–22.

tion Clause jurisprudence.¹²¹ In 1904, Wigmore repeatedly referenced the value and relevance of witness demeanor in his treatise on evidence.¹²² “The demeanor of the witness on the stand may always be considered,” he wrote.¹²³ Wigmore cited cases from the preceding century, but, like those judgments, he neither explained the basis for the demeanor assumption nor explained how judges and juries should actually use demeanor to make decisions. In 1951, the Supreme Court concluded that the value of observing demeanor called for deference not just to an executive agency, but to the actual hearing officer if he or she were to be overruled by superiors.¹²⁴

Is the demeanor assumption correct? The social science about demeanor-based lie detection is controversial, with a divide between researchers who believe it is possible to develop effective techniques and those who argue that no known techniques have been empirically validated. But the researchers on both sides start with a common baseline that has been validated through repeated studies: human beings in general are not usually very good at identifying liars and truth tellers simply by watching and listening to them speak.¹²⁵ In other words, the consensus of researchers is that the assumption long made by our courts about the value of observing demeanor is empirically false. In experiments that ask people to judge whether a speaker is telling the truth or lying by watching and listening to them, respondents are little better than chance at getting it right.¹²⁶

121. 156 U.S. 237, 242–43, 259 (1895); see also Max Minzner, *Detecting Lies Using Demeanor, Bias, and Context*, 29 CARDOZO L. REV. 2557, 2559 (2008) (summarizing the Court’s reliance on demeanor in Confrontation cases from 1895 to the present).

122. 2 ANGLO-AMERICAN SYSTEM OF EVIDENCE, *supra* note 116, § 946 (1904); 3 ANGLO-AMERICAN SYSTEM OF EVIDENCE, *supra* note 116, § 1395.

123. 2 ANGLO-AMERICAN SYSTEM OF EVIDENCE, *supra* note 116, § 946 (emphasis omitted).

124. *Universal Camera Corp. v. NLRB*, 340 U.S. 474, 492–97 (1951).

125. Aldert Vrij et al., *Outsmarting the Liars: Toward a Cognitive Lie Detection Approach*, 20 CURRENT DIRECTIONS IN PSYCHOL. SCI. 28, 28 (2011); Bond & DePaulo, *supra* note 109, at 217 (describing a “consensus” in the research); Glenn Littlepage & Tony Pineault, *Verbal, Facial, and Paralinguistic Cues to the Detection of Truth and Lying*, 4 PERSONALITY & SOC. PSYCHOL. BULL. 461, 463 (1978).

126. See Bond & DePaulo, *supra* note 109, at 216, 219, 230 (discussing previous studies dating more than twenty-five years that consistently found a success rate of 56% or 57%, and their own analysis showing a success rate of 54%, and concluding, “[a]cross hundreds of experiments, typical rates of lie-truth discrimination are slightly above 50%”); Maureen O’Sullivan et al., *The Effect of Comparisons on Detecting Deceit*, 12 J. NONVERBAL BEHAV. 203, 210 (1988). There have been isolated studies reporting a success rate of more than 70%, but a roughly equal number showing success rates much worse than chance, and as low as 31%. But a compilation of results from 292 separate studies found an overall result of between 50% and 60%. Bond & DePaulo, *supra* note 109, at 219. People are slightly more likely to correctly identify an honest speaker (60%) than they are to identify a falsehood (48%). *Id.* at 230.

Perhaps the greatest worry for the legal system is the finding that law-enforcement professionals who have on-the-job experience trying to detect lies do not seem to necessarily gain actual expertise; they are more confident than novices about their accuracy but they are not actually more accurate.¹²⁷ In sum, as Ekman and O'Sullivan wrote in 1991, "it would be important . . . to know how much confidence should be placed in judgments based on demeanor, by layman or expert, about whether someone is lying or telling the truth. The answer from twenty years of research is 'not much.'"¹²⁸

While the general public has little discernible skill at using demeanor to find the truth, there does continue to be scientific controversy about the value of demeanor. The main question is whether there might yet be some prospect of identifying and then teaching more effective lie detection techniques for use in particular contexts. In other words, the open question is whether lie detection might be a specialized skill that may be honed within particular professions. Hope that this may be possible is fueled by two different types of studies. First, Ekman and his colleagues have reported that some people are more accurate than others at identifying truth and falsehoods from demeanor.¹²⁹ If this is true, it might be possible to understand what these people do differently and then to disseminate this knowledge.¹³⁰ Second, close scientific observation and analysis of demeanor has found that there are some reliable cues that correlate to lying.¹³¹ The problem is that most people operate on incorrect assumptions about what these cues are or are not capable of detecting them when they would be useful.¹³²

127. Bella M. DePaulo & Roger L. Pfeiffer, *On-the-Job Experience and Skill at Detecting Deception*, 16 J. APPLIED SOC. PSYCHOL. 249, 249 (1986).

128. Paul Ekman & Maureen O'Sullivan, *Who Can Catch a Liar?*, 47 AM. PSYCHOLOGIST 913, 913 (1991).

129. See *id.*; Paul Ekman et al., *A Few Can Catch a Liar*, 10 PSYCHOL. SCI. 263, 263 (1999).

130. See, e.g., Ekman & O'Sullivan, *supra* note 128, at 919 (citing studies of prior attempts to train observers in identifying "microexpressions" but noting mixed results).

131. See, e.g., Paul Ekman, *Lying and Nonverbal Behavior: Theoretical Issues and New Findings*, 12 J. NONVERBAL BEHAV. 163, 170 (1988) (lying correlated to a rise in vocal pitch); Siegfried L. Sporer & Barbara Schwandt, *Moderators of Nonverbal Indicators of Deception*, 13 PSYCHOL. PUB. POL'Y & L. 1, 2 (2007) (discussing the correlation between physical tension, pupil dilation, and a raised chin and lying).

132. Sporer & Schwandt, *supra* note 131, at 2 (discussing the idea that people errantly believe that blinking, eye contact, gaze aversion, head movements, illustrators, foot and leg movement, nodding, smiling, hand movements, and shifts in posture are all indicators of lying, though research indicates that they are not); Gordon D. Hemsley & Anthony N. Doob, *The Effect of Looking Behavior on Perceptions of a Communicator's Credibility*, 8 J. APPLIED SOC. PSYCHOL. 136, 136 (1978) (finding that speakers who failed to look at the listener directly were

The demeanor signals that Ekman and others describe—raised vocal pitch, pupil dilation, physical tension, and others—appear to correlate with lying, but that does not mean they are directly caused by lying. Instead, they are involuntary indicators of emotional and cognitive strains thought to be related to lying.¹³³ The National Research Council of the National Academy of Sciences has warned that observable physiological signals are most clearly linked to simple emotions but “weak” for more complex mental states, such as deceit.¹³⁴ The same emotions thought to be connected to lying can also result from simple nervousness and anxiety from other causes.¹³⁵ If one assumes that even honest witnesses are often nervous in formal adjudications where a great deal is at stake for them,¹³⁶ the available demeanor signals would be of little utility.¹³⁷ One researcher wrote, “[t]he accumulated evidence suggests that people who are motivated to be believed look deceptive whether or not they are lying.”¹³⁸ Ekman himself warns, “[o]ne must not make *Othello’s error*, of presuming that a sign of fear is a sign of lying.”¹³⁹

Some have argued that a possible solution to this problem is to “make the interview setting more difficult for interviewees” with the goal of making the cognitive demands on liars higher than on truth tellers, on the theory that this will make otherwise ambiguous cues about their honesty more suggestive.¹⁴⁰ This “cognitive lie detection” strategy assumes that telling a lie requires more mental work than telling the truth and thus may involuntarily betray more

more likely to be disbelieved).

133. See Ekman, *supra* note 131, at 165.

134. See STEPHEN M. LORD, AVIATION SECURITY: TSA HAS MADE PROGRESS, BUT ADDITIONAL EFFORTS ARE NEEDED TO IMPROVE SECURITY 7 (2011).

135. An additional problem is that some research suggests that while some facial expressions of emotion are universal, people from different cultures display significantly different capacities to interpret them accurately. Michael Biehl et al., *Matsumoto and Ekman’s Japanese and Caucasian Facial Expressions of Emotion (JACFEE): Reliability Data and Cross-National Differences*, 21 J. NONVERBAL BEHAV. 3, 3 (1997).

136. Cf. 3 ANGLO-AMERICAN SYSTEM OF EVIDENCE, *supra* note 116, § 1395, at 153 n.2 (suggesting that a product of witness confrontation is that it puts the witness under stress).

137. Wigmore himself recognized that it would be an error to infer actual guilt from demeanor, even though he thought it routine to use demeanor to detect lying. ANGLO-AMERICAN SYSTEM OF EVIDENCE, *supra* note 116, § 274.

138. Bond & DePaulo, *supra* note 109, at 231; see also Vrij et al., *supra* note 125, at 28 (“The problem is that cues to deception are typically faint and unreliable. One reason is that the underlying theoretical explanations for why such cues occur—nervousness and cognitive load—also apply to truth tellers. That is, both liars and truth tellers can be afraid of being disbelieved and may have to think hard when providing a statement.” (citation omitted)).

139. Ekman, *supra* note 131, at 167.

140. Vrij et al., *supra* note 125, at 28.

emotion.¹⁴¹ Theoretically, an interviewer can increase the cognitive demand on an interviewee by asking the interviewee to recount a story in reverse chronological order or demanding that the interviewee maintain eye contact, which, in some studies, slightly improved observers' ability to discern truth from falsehood.¹⁴² But, when tested, the improvement achieved through these techniques was slight and may require replication to be validated; in one experiment, lies were correctly identified 60% of the time, more than in the control group,¹⁴³ but only slightly better than the average found in other studies of people's ability to detect lies without the use of any special techniques.¹⁴⁴ Moreover, it is questionable whether raising stress levels produces improvements that would compensate for the obvious downside: impairing the ability of apprehensive but honest witnesses to testify effectively.

Ekman has become the most public promoter of the potential to develop better demeanor-based techniques. His ideas have found recent real-world application in a Transportation Security Administration (TSA) pilot program in Boston in which screeners, who received training in the Screening of Passengers by Observation Techniques (SPOT) program, asked air travelers a few casual questions about their trips.¹⁴⁵ This program expanded an earlier SPOT program that focused on observation of behavior without deliberately engaging passengers in conversation.¹⁴⁶ At least in theory, there may be a logical reason to think that the current state of research has more application in the security screening context than in an adjudication context, assuming that TSA screeners can be trained to do the screening effectively and assuming that the program can be validated through research.¹⁴⁷ In the security screening at an airport, the

141. *Id.* at 28–29.

142. *Id.* at 29.

143. *Id.*

144. *Id.* at 28; see also Aldert Vrij et al., *Increasing Cognitive Load to Facilitate Lie Detection: The Benefit of Recalling an Event in Reverse Order*, 32 *LAW & HUM. BEHAV.* 253, 253–65 (2008) (finding lies correctly identified 53% of time when liars were instructed to tell their stories in reverse chronological order); Aldert Vrij et al., *'Look into My Eyes': Can an Instruction to Maintain Eye Contact Facilitate Lie Detection?*, 16 *PSYCHOL. CRIME & L.* 327, 327–48 (2010) (finding lies correctly identified 54% of the time when liars were instructed to maintain eye contact).

145. Andrew Seidman, *TSA Launching Behavior-Detection Program at Boston Airport*, *L.A. TIMES* Aug. 17, 2011, <http://articles.latimes.com/2011/aug/17/nation/la-na-tsa-logan-20110818>; LORD, *supra* note 134.

146. LORD, *supra* note 134, at 7–11.

147. *But see* Michael S. Schmidt & Eric Lichtblau, *Racial Profiling Rife at Airport, U.S. Officers Say*, *N.Y. TIMES* (Aug. 11, 2012), <http://www.nytimes.com/2012/08/12/US/racial-profiling-at-boston-airport-officials-say.html> (more than thirty TSA officers criticizing the Boston pilot

purpose of the demeanor observation is not to reach a final conclusion that the subject is dangerous, but rather to identify passengers who should receive further scrutiny. In 2010, the SPOT pilot program referred 50,000 passengers for additional screening, and referred 3600 to law enforcement authorities.¹⁴⁸ These referrals produced 300 actual arrests.¹⁴⁹ This might be considered a relative success at identifying high-risk passengers at an airport,¹⁵⁰ but it suggests that more than 99% of those people initially flagged by observation of demeanor were ultimately innocent. That would be a disastrous error rate in any adjudication context, and reflects a general challenge to applying current research about demeanor and deceit in adjudication.

This is an active area of research by both the government and academic scholars. It is of course possible that eventually there will be a validated technique by which to discern truth from lies through demeanor observation. But, because such an achievement would be so attractive for so many reasons, we should guard against wishful thinking. Charles F. Bond and Bella M. DePaulo offer this warning in a widely cited study:

We see a pattern in this research literature. In their reading of the literature, scholars find an unwanted implication—that people can barely discriminate lies from truths. Heirs to the moralistic tradition, scholars resist this implication by identifying a feature of researchers' methods that could in principle explain low lie-truth discrimination rates. . . . Rather than marveling at outliers in this literature, we are more impressed by the regularity of the results obtained. . . . We wonder if it is premature to abort the quest for 90% lie detection and accept the conclusion implied by the first 384 research samples—that to people who must judge deception in real time with no special aids, many lies are undetectable.¹⁵¹

program for being a mask for racial and ethnic profiling). The General Accounting Office has repeatedly raised concerns that there is insufficient empirical research to adequately validate the SPOT program, and has recommended convening an expert panel to attempt to do so. LORD, *supra* note 134, at 7, 9.

148. LORD, *supra* note 134, at 7.

149. *Id.*

150. *See id.* at 9 ("SPOT was more effective than random screening . . . at identifying individuals who possessed fraudulent documents and identifying individuals who law enforcement officers ultimately arrested.").

151. Bond & DePaulo, *supra* note 109, at 231.

Doubts about the utility of demeanor could lead to the false impression that it is simply impossible to adjudicate facts reliably. Despite all the doubt about demeanor, social science has pointed to approaches to witness credibility that do in fact show substantial effectiveness.¹⁵² People telling falsehoods tend to be less logical and less consistent.¹⁵³ Researchers have found success (80% correct identification of deceit) by asking potential liars for specific descriptions of events, especially by requesting details that the interviewee might not have anticipated before the interview.¹⁵⁴ These experiments have found significant differences in the level of detail provided by truth-tellers and liars when asked to describe, through words and sketches, a physical location or a personal opinion, liars gives less detail.¹⁵⁵ Similarly, interviewers can correctly identify lies with 85% accuracy when they can strategically ask questions and compare the answers to a piece of external evidence unknown to the interviewee.¹⁵⁶ Yet, the primary cue to deceit in these experiments was the inconsistent *content* of answers, not nonverbal signals.¹⁵⁷ Because such credibility assessment strategies rely on matters easily captured in a written record, they do not provide justification for an appellate court to defer to a first instance decision maker.¹⁵⁸

For the American legal system, the unsettling headline is that a heavy pile of empirical research directly rebuts the long-held assumption that people can detect truth and lies when they see a witness face-to-face. While the search continues for a way to use demeanor observation for law enforcement, prospects for doing so are limited and there is reason for skepticism about their effectiveness even in narrowly defined contexts. In addition, there are separate bodies of empirical research suggesting that live observation of a witness can actually make a decision maker more inaccurate be-

152. See, e.g., Vrij et al., *supra* note 125, at 29–31; Minzner, *supra* note 121, at 2567–71.

153. Minzner, *supra* note 121, at 2569.

154. Vrij et al., *supra* note 125, at 29.

155. *Id.* at 30.

156. Minzner, *supra* note 121, at 2569.

157. *Id.*

158. See *Kadia v. Gonzales*, 501 F.3d 817, 819 (7th Cir. 2007) (“In a case such as this, in which the basis for the evaluation of the witness’s credibility is set forth in detail by the trier of fact and has nothing to do with demeanor but consists instead of inconsistencies or falsehoods in the witness’s testimony that the trier of fact enumerates in his opinion, the reviewing court has more than suspicion to work with in deciding whether the determination of credibility was reasonable.”); *Gao v. Bd. of Immigration Appeals*, 482 F.3d 122, 127 (2d Cir. 2007) (holding that less deference is granted to credibility determinations not based on demeanor).

cause people are more likely to believe witnesses who are more physically attractive, more similar to themselves, or appear to have high social status by virtue of race, gender, clothing, grooming, and manner of speech.¹⁵⁹ One study found that people are simply more likely to believe extroverted and socially skilled speakers, and less likely to believe speakers who are “socially anxious,” regardless of whether they are actually telling the truth.¹⁶⁰ The obvious but unsettling conclusion is that, on the question of assessment of witness credibility, our legal tradition has been making an unfounded assumption for centuries. Not only is there little reason to think that first-instance adjudicators can judge witness credibility better than appellate judges, there is good empirical reason to conclude that they cannot.

Despite all of this available knowledge, at the dawn of the twenty-first century the Supreme Court suggested, without citation, that assessment of witness credibility “is not susceptible to much further refinement.”¹⁶¹ We thus face two different challenges in asking questions about the demeanor assumption. First, now that social science has built up a body of empirical knowledge about human lie detection capacities, it would seem logical to begin to ask whether the law’s empirical assumption has any validity. And, yet, because the law has for so long assumed that demeanor works, the law actually resists this inquiry. We thus confront a primary assumption—people can judge the truth through demeanor—that is defended from empirical attack by a secondary assumption that it would be futile to try to improve human lie detection. Our legal tradition has thus used an assumption about how the world works as a foundation of our system of justice and simultaneously insulated this assumption from expanding empirical knowledge about how the world actually works. Like other empirical research that questions the accuracy and biases of human beings as witnesses and interpreters of events, this finding should call for considerably more caution

159. Spottswood, *supra* note 107, at 844–46.

160. See generally Ronald E. Riggio et al., *Social Skills and Deception Ability*, 13 PERSONALITY & SOC. PSYCHOL. BULL. 568 (1987) (discussing researchers’ investigation of individuals’ abilities to deceive using video-taped messages).

161. *Bush v. Gore*, 531 U.S. 98, 106 (2000) (explaining why a uniform standard should exist to interpret physical evidence, such as ballot cards, but would not be required for interpretation of witness testimony); cf. Bond & DePaulo, *supra* note 109, at 214 (noting that the legal system avoids the complex challenge posed by deception by deferring the matter to juries who are advised to judge credibility based on demeanor and are presumed to be good at lie detection).

by our legal system about the capacities of its decision makers.¹⁶²

One reason for the persistence of the demeanor assumption is that it appeals to a romantic notion of how truth will emerge through the drama of a courtroom trial.¹⁶³ As recently as 1988, the Supreme Court explained the value of witness confrontation in a criminal trial by asserting that a false accuser would avoid eye contact with the defendant.¹⁶⁴ That case, *Coy v. Iowa*, struck down a state law that allowed victims in a sex abuse case to testify with a screen shielding them from the defendant, but not from the jury.¹⁶⁵ Justice Scalia's majority opinion made reference to "something deep in human nature,"¹⁶⁶ but cited no scientific evidence. Instead, it quoted President Eisenhower discussing the importance of airing disputes face-to-face in the traditional value system of his hometown of Abilene¹⁶⁷ and noted the continued use of the phrase "[l]ook me in the eye" in every-day English.¹⁶⁸ The Court was satisfied to base its decision on an empirical assumption evidenced by little more than the Court's confident assertions of the assumption "[t]he perception that [face-to-face] confrontation is essential to fairness has persisted over the centuries because there is much truth to it."¹⁶⁹ That circular form of reasoning may explain succinctly the persistence of the demeanor assumption. Courts take a central foundation for a major tenet of American law to be true simply because earlier courts endorsed it, producing an insular and circular body of legal thought divorced from knowledge.¹⁷⁰ At some point, one hopes, courts will begin to question whether this is how our legal system should operate.

162. See Dan M. Kahan et al., "They Saw a Protest": Cognitive Illiberalism and the Speech-Conduct Distinction, 64 STAN. L. REV. (forthcoming 2012) (calling for "judicial humility" as a response to evidence of cognitive bias).

163. Justice Scalia's majority opinion in *Coy* captures this romantic approach, quoting self-consciously from "references to and quotations from antiquity," coupled with the common sense wisdom of the Midwest. *Coy v. Iowa*, 487 U.S. 1012, 1017 (1988).

164. See *id.* at 1019-20.

165. *Id.* at 1022.

166. *Id.* at 1017.

167. See *id.* at 1017-18.

168. *Id.* at 1018.

169. *Id.* at 1019. The dissent did not dispute these assumptions directly but instead noted that the screen did not prevent the jury from viewing the child witness' demeanor, and that the court had lost its focus on the central value of cross-examination, rather than eye contact. See *id.* at 1027-28 (Blackmun, J., dissenting).

170. For a discussion of how *stare decisis* can sometimes create errors of logic, see Goutam U. Jois, *Stare Decisis Is Cognitive Error*, 75 BROOK. L. REV. 63 (2009).

B. *The Comparative Role of Juries, Judges, and Agencies*

Although this Article is about administrative adjudication in the immigration arena, the widespread reliance on the demeanor assumption in our system of justice requires some comment, especially as it relates to the role of juries. This Article should not be read as an attack on juries *per se*, though the concern about demeanor is certainly relevant to juries. Juries are a unique institution. More to the point, it is dangerous to assume that legal norms adapted—and perhaps justified—in the case of juries can be transplanted into administrative adjudication without additional justification.

Without question, the disappointing headline about people's general inability to detect lies may be damning of juries because it is through juries that the legal system most clearly utilizes the public's supposed ability to judge the truth. Many of the rationales for appellate deference to findings of facts developed first in reference to jury trials, and doubts that we have seen about the actual utility of viewing witnesses in person would certainly raise doubts about juries' capacity to fulfill this role as well.¹⁷¹ Does this mean that the strong constitutional protection of a right to a jury trial is misplaced? The Bill of Rights guarantees a jury trial in federal criminal cases,¹⁷² a right extended through the Fourteenth Amendment to state criminal defendants.¹⁷³ The Seventh Amendment also guarantees jury trials in civil cases in federal court,¹⁷⁴ explicitly providing that "no fact tried by a jury, shall be otherwise reexamined in any Court of the United States, than according to the rules of the common law."¹⁷⁵ Many state constitutions also guarantee jury trials in civil cases.¹⁷⁶

The empirical research I have discussed here suggests the assumption that juries have a special ability to make accurate judgments of facts may not be the most sound foundation on which to justify the American jury system.¹⁷⁷ Yet, there are other reasons to

171. See RANDOLPH N. JONAKAIT, *THE AMERICAN JURY SYSTEM* 52 (2003) (noting that "jurors are amateurs" and are not better than judges at assessing witness credibility).

172. U.S. CONST. amend. VI.

173. *Duncan v. Louisiana*, 391 U.S. 145, 156 (1968).

174. U.S. CONST. amend. VII ("In Suits at common law, where the value in controversy shall exceed twenty dollars, the right of trial by jury shall be preserved.").

175. *Id.*

176. See, e.g., CAL. CONST. art. I, § 16; N.Y. CONST. art. I, § 2.

177. *But see* *United States v. Scheffer*, 523 U.S. 303, 313 (1998) (arguing that, among other reasons, polygraph tests could be excluded from criminal trials because they would "diminish the jury's role in making credibility determinations" and that "[a] fundamental premise of our criminal trial system is that the jury is the lie detector" (internal quotation marks omitted));

favor juries. First and foremost, juries exist “to prevent oppression by the Government . . . to protect against unfounded criminal charges brought to eliminate enemies and against judges too responsive to the voice of higher authority.”¹⁷⁸ This is not only a reason for juries to exist, but also to restrain the degree to which appellate judges can overrule jury decisions. This weighty concern about the role of juries in American democracy deserves far more extensive treatment than I have provided here. But it is important in looking for a rationale for the role of juries, to focus on functions that juries are uniquely well-placed to fulfill. Absent new data, the ability to judge witness credibility would not appear to be such a function. By contrast, juries may be well-placed to determine community values and norms of conduct in cases where such assessments are critical.¹⁷⁹

While there are unique reasons to protect the role of juries as finders of fact, the special place of juries in our legal system can cause confusion when their functions are too quickly transposed to administrative agencies. If juries exist as a safeguard against government tyranny, it would be odd, indeed, to think that an executive agency could fulfill a similar function. But consider a view offered by then-Judge Alito in an asylum appeal handled by the Third Circuit. Arguing in a dissent that the court should defer to factual inferences made by the Immigration Judge, Judge Alito emphasized the concept of “background knowledge,”¹⁸⁰ which Judge Weinstein of the Eastern District of New York had defined as the “vast storehouses of commonly-held notions about how people . . . generally behave.”¹⁸¹ Judge Alito argued, “this process of judging credibility is employed every day in criminal and civil cases, and there is no reason why it cannot also be used in asylum cases.”¹⁸²

Curiously, Judge Alito quoted from an article discussing how jurors determine whether witness testimony is plausible, including a passage warning that such inferences are difficult when a witness comes from a different culture,¹⁸³ which is the norm in asylum cases. After providing this warning, Judge Alito neglected to fully explain

United States v. Lumpkin, 192 F.3d 280, 289 (2d Cir. 1999) (stating the fundamental role of the jury is to assess witness credibility).

178. *Duncan*, 391 U.S. at 155–56.

179. See JONAKAIT, *supra* note 171, at 64–66, 71–72.

180. *Dia v. Ashcroft*, 353 F.3d 228, 262 (3d Cir. 2003).

181. *United States v. Shonubi*, 895 F. Supp. 460, 479 (E.D.N.Y. 1995), *vacated*, 103 F.3d 1085 (2d Cir. 1997).

182. *Dia*, 353 F.3d at 263.

183. See *id.* at 263 n.2.

why this warning should not be heeded except to say an immigration judge, much like a civil or criminal jury, may make a credibility decision “by comparing that testimony to the IJ’s background knowledge about human behavior in general and about the behavior of those seeking entry into the United States.”¹⁸⁴ The question raised by Judge Alito’s argument is whether an immigration judge, or, for that matter, any administrative adjudicator, is really equivalent to a jury *for the purpose of appellate review* of inferences about human behavior.

Let us assume, *arguendo*, that juries are well-suited to draw inferences about general norms of behavior because jurors are drawn from the community and, thus, know well (or at least better than judges) how community members behave and that this knowledge would somehow be helpful in assessing the behavior of an immigrant to the community. This assumption would then justify judges deferring to jurors as they make these inferences. But, when there is no jury, the question is whether an appellate panel of judges should defer to a lower judge or an administrative adjudicator when he or she makes such inferences. It is hard to see why an administrative adjudicator is *per se* a better judge of human behavior than a court of appeals judge, except for the possibility that particular adjudicators develop particular areas of expertise,¹⁸⁵ a suggestion that I will address in Part V, Section A. The bottom line is that juries are a unique institution, and what is true of them cannot automatically be assumed to be true of other decision makers.

III. THE EFFICIENCY RATIONALE

A. *An Independent Rationale for Deference?*

In the 1932 worker’s compensation case *Crowell v. Benson*, the Supreme Court held that an administrator’s findings of fact, if supported by evidence, should be final so as to not “defeat the obvious purpose of the legislation to furnish a prompt, continuous, expert, and inexpensive method for dealing with a class of questions of fact which are peculiarly suited to examination and determination by an administrative agency specially assigned to that task.”¹⁸⁶ More re-

184. *Id.* at 263.

185. See *Anderson v. City of Bessemer*, 470 U.S. 564, 574 (1985) (“The trial judge’s major role is the determination of fact, and with experience in fulfilling that role comes expertise.”).

186. 285 U.S. 22, 46 (1932).

cently, in *Anderson v. City of Bessemer*, the Court asserted that judicial efficiency is an independent justification for giving deference to findings of fact.¹⁸⁷ The independent nature of the efficiency rationale was critical for the court in order to explain why deference should be granted even when a factual finding was not dependent on assessment of witness credibility.¹⁸⁸ The Court said:

Duplication of the trial judge's efforts in the court of appeals would very likely contribute only negligibly to the accuracy of fact determination at a huge cost in diversion of judicial resources. In addition, the parties to a case on appeal have already been forced to concentrate their energies and resources on persuading the trial judge that their account of the facts is the correct one; requiring them to persuade three more judges at the appellate level is requiring too much. As the Court has stated in a different context, the trial on the merits should be "the 'main event' . . . rather than a 'tryout on the road.'"¹⁸⁹

Despite the Court's assertions, I would suggest that it is still an open question whether efficiency is a fully independent rationale for deference. Because courts continue to reassert the assumption that first-instance adjudicators are more accurate, they rarely have to rely on the efficiency rationale as a stand-alone justification for deference. In *Anderson*, the Court hedged by claiming without explanation that stronger appellate review would "only negligibly" improve factual accuracy.¹⁹⁰ This reflects a perception by the Court that efficiency in adjudication can be achieved with no real costs in terms of accuracy. If we assume that a first-instance decision is the one most likely to be accurate anyway, then it would seem that nothing would be gained by close appellate review. If we assume accuracy, then the persuasive burden shifts to the party that wants to appeal who must explain why more judicial resources should be spent on a case already presumed to be decided correctly. Efficiency in this view is a secondary benefit of accuracy.

The way the Court linked accuracy and efficiency in *Anderson* contrasts with the way the Court recently discussed the value of efficiency in an arbitration context, where it acknowledged that effi-

187. See *Anderson*, 470 U.S. at 574-75.

188. *Id.* at 574.

189. *Id.* at 574-75 (citing *Wainwright v. Sykes*, 433 U.S. 72, 90 (1977)).

190. *Id.* at 575.

ciency comes at a price. Explaining why it is important to keep the stakes of arbitration low, the Court said:

Informal procedures do of course have a cost: The absence of multilayered review makes it more likely that errors will go uncorrected. Defendants are willing to accept the costs of these errors in arbitration, since their impact is limited to the size of individual disputes, and presumably outweighed by savings from avoiding the courts. But when damages allegedly owed to tens of thousands of potential claimants are aggregated and decided at once, the risk of an error will often become unacceptable.¹⁹¹

Given the empirical reasons to doubt the automatic accuracy of first-instance adjudicators, courts should be cautious about assuming that efficiency comes at no real cost. Efficiency and accuracy are often in competition, and trade-offs must be made. This does not mean that efficiency should be abandoned, but the reality of a trade-off should shape the way courts seek to achieve it. Efficiency is desirable, no doubt, but only when it is understood to come at a cost can we see how courts actually prioritize its value.

B. Streamlining and Deterrence

To understand the potential connection between judicial efficiency and appellate standards of review, it is important to be precise about what kind of efficiency we hope to achieve. Efficiency calls out for a balance. After all, the simplest and most direct way to achieve judicial efficiency would be to eliminate judicial processes entirely. This is done in administrative agency actions regularly, whenever an agency decision is not open to appeal even at the administrative level.¹⁹² But to do this in all areas of adjudication would be objectionable because of the value of due process.

One way judicial efficiency might be achieved through deference is to streamline or shorten the work of an appellate court and to prevent a complete re-trying of the case. This was the primary con-

191. *AT&T Mobility v. Concepcion*, 131 S. Ct. 1740, 1752 (2011).

192. For examples of immigration case types not subject to judicial review, see, for example, Memorandum from U.S. Dep't of Homeland Security on Implementation of the Help HAITI Act of 2010 (March 13, 2011) available at http://www.uscis.gov/USCIS/Laws/Memoranda/2011/April/Help_HAITI_Act_PM.pdf ("There is no administrative appeal from a denial of a Help HAITI act application.").

cern expressed by the Supreme Court in *Anderson*.¹⁹³ A first-instance adjudicator who produces a written decision must receive, sort, and analyze an evidentiary record in the course of producing findings of fact. It is understandable that duplicating this effort on appeal would be inefficient, but it is unclear how deference achieves this purpose. There are other rules limiting the right of parties to introduce new evidence on appeal and requiring them to base their assertions on the record.¹⁹⁴ In other areas of law, the Supreme Court has embraced limitations on the jurisdiction of federal courts to conduct evidentiary hearings in order to “channel” claims into the state courts.¹⁹⁵ The BIA itself is not permitted to reach new findings of fact in considering an appeal from an immigration judge. But these are procedural rules entirely separate from the standard of review.¹⁹⁶

In theory, the substantial evidence or clearly erroneous standard allows an appellate judge to read the lower decision, see that there is a rational basis cited to the evidential record, and reject the appeal. But, once an appellant argues that the first-instance adjudicator misconstrued a piece of evidence or took it out of context, it may be difficult for a thorough appellate judge to avoid a deeper parsing of the evidentiary record. Moreover, the practice of remanding cases for further fact-finding when error is found on appeal creates its own inefficiency because it leaves cases bouncing up and down the levels of adjudication.¹⁹⁷

The nature of the deference doctrine suggests that appellate judg-

193. See *Anderson*, 470 U.S. at 573.

194. See, e.g., 8 U.S.C.A. § 1252(b)(4)(A) (West 2005) (authorizing the court of appeals to review removal orders “only on the administrative record on which the order of removal is based”); see also Hiroshi Motomura, *Judicial Review in Immigration Cases After AADC: Lessons from Civil Procedure*, 14 GEO. IMMIGR. L.J. 385, 444–45 (2000). The Federal Rules of Appellate Procedure (FRAP) require appellants in administrative cases to “specify the order or part thereof to be reviewed.” FED. R. APP. P. 15(a)(2)(C). The FRAP also specify that the record on appeal consists of the agency’s order, its findings or report, and the pleadings, evidence, and proceedings before the agency. FED. R. APP. P. 16. However, parties on appeal may supplement the record. *Id.*

195. *Cullen v. Pinholster*, 131 S. Ct. 1388, 1398–99 (2011) (limiting federal habeas corpus review to the evidence presented in the state court).

196. The procedural rule that first-instance adjudicators have primary responsibility for factual adjudication can build a form of inefficiency, in that it leads appellate courts to remand for re-consideration rather than to have all matters decided finally on appeal.

197. *Fla. Power & Light Co. v. Lorion*, 470 U.S. 729, 744 (1985) (“If the record before the agency does not support the agency action, if the agency has not considered all relevant factors, or if the reviewing court simply cannot evaluate the challenged agency action on the basis of the record before it, the proper course, except in rare circumstances, is to remand to the agency for additional investigation or explanation.”).

es will be able to apply a more cursory analysis, but its real benefit may be something else. Even if appellate judges engage in deep analysis of an evidentiary record, appellate standards of review may yet achieve a kind of streamlining if they help judges keep this analysis narrowly focused. What passes for deference at the surface may at a more functional level be a demand that appellants point specifically to the errors they believe were made at first instance rather than ask for a total review of the case from the beginning. This then allows an appellate court to focus on specific aspects of a lower decision and often on only specific pieces of evidence. Even if the appellate court decides the issue on a *de novo* basis, based on what it thinks is most likely correct, it has still been able to skip over the vast amount of evidentiary analysis that the first-instance adjudicator would have done. It is the difference between doing a home renovation and demolishing the house and building from scratch.¹⁹⁸

Another way courts may achieve efficiency is to discourage the filing of appeals.¹⁹⁹ For this to occur, the party who has lost at first instance—typically a claimant denied some benefit by an agency—must perceive that it would not be worthwhile to try to change the outcome. For this to be a rational choice, three factors must come into play. First, appellants must perceive a relatively high transaction cost in pursuing the appeal. Second, they must perceive a low chance of success on appeal. Third, they are more likely to decline the opportunity to appeal if they do not see that a great deal is at stake anyway. When one thinks about how these factors compute, it should be clear this type of efficiency is more achievable in some types of cases than in others.

The clearest manifestation of this principle is the relationship between the interest at stake and the chance of success. Many resources and interests are in play in decisions to appeal, including money and assets, emotional stakes, matters of principle (and perceived principle), time, possibly even deportation, and bodily integrity (in an asylum case). But, for illustration, it may be possible to

198. Linda Berger usefully illustrates this as a camera lens that becomes more narrow as a case moves through stages of litigation. Before trial, lawyers and clients have the widest angle of focus, encompassing the entire universe of circumstances related to the case. But as they make strategic decisions about what evidence to present, and as this is filtered through the rules of evidence, cross-examination, argument and then analysis by the adjudicator, the camera lens begins to narrow. On appeal, the lens is focused on specific decisions by the adjudicator. Even if the standard of review is *de novo*, appeal is not an opportunity to return to the full universe of available evidence.

199. See Cooper, *supra* note 83, at 646.

reduce these decisions to a dollar-value comparison. Assume the cost of pursuing an appeal is \$10,000 for a hypothetical appellant.²⁰⁰ Assume also that the appellant anticipates a 20% chance of prevailing, and has interests valued at \$100,000 to gain if she does prevail. This means the real value of the appeal for the appellant is \$20,000, and she has good reason to pursue the appeal. But if her chances of prevailing were reduced to, say, 5% or if the value of her interests in the case was reduced to \$40,000, she would have good reason not to appeal. Thus, a lower anticipated chance of success for potential appellants will achieve greater efficiency only when the interests at stake are low.

A COST OF APPEALING	B LIKELIHOOD OF PREVAILING	C INTEREST AT STAKE	D REAL VALUE OF THE APPEAL (B x C)	LIKELY ACTION (COMPARE A AND D)
\$10,000	20%	\$100,000	\$20,000	Appeal
\$10,000	5%	\$100,000	\$5000	No Appeal
\$10,000	20%	\$40,000	\$8000	No Appeal
\$10,000	1%	Life or Liberty	Priceless	Appeal

But there is another reason why the efficiency rationale is most likely to be viable when the interests at stake are small. If judicial efficiency is to be achieved at the expense of accurate decision-making, this tradeoff will be much harder to justify when the stakes are high. Moreover, designing a system to deter appeals when there are substantial interests at stake may be antithetical to values of access to justice in a democracy.²⁰¹ In an asylum context, if a person

200. These calculations are for illustration only, and translate values to numerical figures that do not have any objective value. Moreover, these calculations are based on the appellant's perceptions.

201. For illustration, I have used a figure of \$10,000 as the cost of appealing. This figure is intended to include the actual monetary costs—lawyers, filing fees, etc.—as well as the emotional and other less directly quantifiable strains of prolonging litigation. In practice, many people file appeals *pro se*, and filing fees can often be waived for the indigent.

genuinely believes his life to be in danger, it would be rational to appeal even with a 1% chance of success. Moreover, many appellate judges may find it difficult to put aside doubts about the accuracy of a decision when the consequences of error would be so grave.²⁰² This may be why streamlining the Board of Immigration Appeals did not pay clear dividends in terms of efficiency.

If the deference doctrine is to achieve real gains in terms of efficiency, there is good reason to try to reform the doctrine itself. As we have seen, a generic administrative law appellant today will win against an agency at the court of appeals level roughly one third of the time. Although it should be noted that many of these “wins” are remands where the final outcome remains in doubt, this rate of appellate success may not be low enough to deter many parties. Even more important may be the fact that it has been difficult to find empirical differences between de novo and deference appeals. This raises doubts about whether deference actually lowers chances of success at all chiefly because the deference doctrine appears to be inconsistently applied and subject to ambiguity. Appellants can be deterred only if they perceive a predictable low chance of success before filing. The judicial flexibility—and thus inconsistency—that has been built into deference may thus work against one of the doctrine’s ultimate goals.

IV. DELEGATION OF POWERS

The Constitution assigns “all legislative powers” to Congress,²⁰³ and Congress generally cannot delegate its legislative authority to another branch of government.²⁰⁴ But, in the terminology used by the courts, Congress may seek the assistance of another branch of government so long as it sets out an “intelligible principle to which the person or body authorized . . . is directed to conform.”²⁰⁵ The Supreme Court has explained this allowance for delegation by Congress’s need to develop laws that fit a complex and technical world.²⁰⁶ By the Court’s own account, it has permitted “broad dele-

202. See *infra* Part IV.

203. U.S. CONST. art. I, § 1.

204. *Mistretta v. United States*, 488 U.S. 361, 372 (1989); *Marshall Field & Co. v. Clark*, 143 U.S. 649, 692 (1892).

205. *Mistretta*, 488 U.S. at 372 (adopting the approach found in *J.W. Hampton, Jr., & Co. v. United States*, 276 U.S. 394, 406 (1928)).

206. See, e.g., *id.*; *Whitman v. Am. Trucking Ass’ns*, 531 U.S. 457, 472 (2001) (permitting the Environmental Protection Agency to set clean air standards); *Touby v. United States*, 500 U.S.

gations" by Congress,²⁰⁷ except in two cases in 1935 where the Court struck down New Deal-era trade regulations.²⁰⁸ The Court's willingness to permit liberal delegation by Congress has been the subject of academic debate.²⁰⁹ Some commentators argue that unrestrained delegation thwarts democratic accountability for rule-making because it allows Congress to pass on the burden of writing rules that might prove unpopular.²¹⁰ Others, including now-Justice Kagan, have argued that the regulatory state is a central Presidential tool for pursuing policy agendas and is thus subject to political accountability through elections.²¹¹

Yet, the main debate about delegation jurisprudence focuses on rule making, an issue more linked to *Chevron* deference than to the adjudication of facts in individual cases. It may be easier to find a democratic rationale for delegating rule making to an executive agency than to delegate individual case adjudication. For rule making, it may be easier to bring public pressure and accountability to bear on the Executive Branch.²¹² Appellate deference to administrative adjudication raises a different question, especially if the deference is directed by Congress. In short, should the courts allow the legislature to transfer adjudication authority from the judiciary to the executive?²¹³ Given that Article III confers the "judicial power of

160, 164–67 (1991) (permitting Congress to delegate to the Attorney General authority to add drugs to the controlled substances list).

207. *Mistretta*, 488 U.S. at 374.

208. See generally *A.L.A. Schechter Poultry Corp. v. United States*, 295 U.S. 495 (1935); *Panama Refining Co. v. Ryan*, 293 U.S. 388 (1935).

209. See David Horton, *Arbitration as Delegation*, 86 N.Y.U. L. REV. 437, 470–71 (2011) (summarizing the major arguments).

210. DAVID SCHOENBROD, *POWER WITHOUT RESPONSIBILITY: HOW CONGRESS ABUSES THE PEOPLE THROUGH DELEGATION* 10 (1993).

211. See Elena Kagan, *Presidential Administration*, 114 HARV. L. REV. 2245, 2369 (2001) ("The President's involvement, at least if publicly disclosed, vests the action with an increased dose of accountability, which although not (by definition) peculiarly legislative in nature, renders the action less troublesome than solely bureaucratic measures from the standpoint of democratic values.").

212. The Supreme Court recently endorsed the rule of executive discretion to accommodate the "immediate human concerns" that might otherwise be implicated by strict application of immigration law. See *Arizona v. United States*, 132 S. Ct. 2492, 2499 (2012). Such discretion would appear consistent with Kagan's theory, especially when it is announced by the President in a high profile manner, as has recently been the case in immigration enforcement.

213. See *N. Pipeline Constr. Co. v. Marathon Pipe Line Co.*, 458 U.S. 50, 76 (1982) (plurality holding that Article III "bars Congress from establishing legislative courts to exercise jurisdiction over all matters related to those arising under the bankruptcy laws"); Merrill, *supra* note 55, at 996 (describing judicial anxiety about administrative encroachment into the judicial domain).

the United States” on the judiciary, it is surprising how readily our legal system has accepted the expansion of non-judicial administrative adjudications that would appear to be exercising judicial power.²¹⁴ In this, much as in the deference doctrine, the Court and Congress have used categorical rules that are subject to considerable ambiguity.

The prevailing explanation for the way the Supreme Court has understood delegation of Article III powers in this context was expressed by Justice Brennan for a plurality of the Court in *Northern Pipeline Construction Co. v. Marathon Pipe Line Co.*, where he wrote that Congress may establish special tribunals to adjudicate “public rights” between the government and private parties.²¹⁵ By contrast, “private rights” that fall under traditional common law domains, such as contracts and torts, would remain tightly held by the judiciary.²¹⁶ Yet, the public rights doctrine has difficulty explaining the early twentieth-century railroad cases that dealt with private economic relationships,²¹⁷ as well as later cases that also concerned “private rights” but where the court nevertheless approved a delegation of power. In a Depression-era employment law case, the Court cited the public rights doctrine to explain Congress’s prerogative to delegate factual adjudication to administrative agencies, specifically citing “interstate and foreign commerce, taxation, immigration, the public lands, public health, the facilities of the post office, pensions, and payments to veterans.”²¹⁸ But, then noting that it was deciding a case concerning employment in the private sector outside the public rights category, the Court went on to say that, even in a private rights case, “there is no requirement [under Article III] that, in order to maintain the essential attributes of the judicial power, all determinations of fact in constitutional courts shall be made by judges.”²¹⁹ Thus, though the Court has repeatedly cited the public-private distinction, it is not clear that this distinction actually explains the outcome of the cases.

The public rights doctrine is awkwardly rooted in the pre-Civil War era when sovereign immunity largely blocked citizens from challenging official actions in court based on a doctrine with origins

214. Merrill, *supra* note 55, at 943, 979–80.

215. *N. Pipeline Constr. Co.*, 458 U.S. at 69–70.

216. *Id.*

217. Merrill, *supra* note 55, at 986–87.

218. *Crowell v. Benson*, 285 U.S. 22, 51 (1932).

219. *Id.*

in the tradition of sovereign immunity.²²⁰ One could reasonably ask whether it makes sense to apply such a tradition to the relationship between citizens and government agencies today. Merrill argues that the lack of a coherent explanation for the delegation of powers is a product of the political history that produced the relevant cases. As the Court retreated from its early assertive interventions into railroad regulation, it feared being driven entirely from the increasingly important fields of administrative law. Much like the ambiguities of the deference doctrine more generally, the fuzzy lines between public and private rights provided a flexible mechanism to restrain the judiciary from ranging too far into the realm of executive administration, while preserving the courts' ability to defend their own ultimate mandate to determine questions of law.²²¹

When Congress delegates adjudication to an agency, judicial deference to the agency extends only to those matters "within the scope of [its] authority."²²² This focus of administrative law connects with what legal philosophers have called a "culture of authority," where the legitimacy of a government action focuses on whether the actor was authorized to act, but avoids scrutiny of what the actor actually did with this authority.²²³ But, even as the Court focused on questions of executive authority, it made clear that the courts would define the boundaries of that authority. Here, we can see an interesting interaction of the delegation of powers doctrine with the fact-law distinction that lies at the head of judicial deference in adjudication. On the one hand, a court may defer to an agency that Congress has authorized to adjudicate a certain type of case. But the courts must interpret the scope of the Congressional authorization, which is a question of law not subject to deference. As the Court wrote in *Crowell*:

Congress did not attempt to define questions of law, and the generality of the description leaves no doubt of the intention to reserve to the Federal court full authority to pass upon all matters which this Court had held to fall within that category. There is thus no attempt to interfere with, but rather provision is made to facilitate, the exercise by the court of its jurisdiction to deny effect to any administrative

220. *Murray's Lessee v. Hoboken Land & Improvement Co.*, 59 U.S. 272, 284 (1855).

221. See Merrill, *supra* note 55, at 997-98.

222. *Crowell*, 285 U.S. at 46.

223. See Moshe Cohen-Eliya & Iddo Porat, *Proportionality and the Culture of Justification*, 59 AM. J. COMP. L. 463, 475 (2011).

finding which is without evidence, or “contrary to the indisputable character of the evidence,” or where the hearing is “inadequate,” or “unfair,” or arbitrary in any respect.²²⁴

Again, we see how a doctrine that appears to limit judicial involvement in effect offers, through ambiguous categorical distinctions, significant judicial prerogatives to intervene when courts want to intervene as well as rationales for judicial restraint when courts want to restrain themselves. As Merrill explains:

Insofar as the reviewing institution wants to overturn an agency decision on an issue within the sphere of competence of the agency, it can usually find a way to do so that suggests it is exercising its own competence. Thus, for example, the reviewing institution will overturn a fact-based decision by the agency by describing the decision as so lacking in evidence as to be “contrary to law.” Alternatively, insofar as a court wants to uphold the decision of an agency, it will frame the issue in terms of the competence of the agency, for instance by positing that whether a carrier’s practice is discriminatory is a “question of fact” for the agency to determine in its “sound discretion.”²²⁵

In immigration, this tendency can be seen in section 242 of the Immigration and Nationality Act, which authorizes judicial review by the courts of appeal.²²⁶ Congress has prohibited the federal courts from reviewing immigration decisions that are discretionary or removal orders against noncitizens with criminal records.²²⁷ Yet, Congress acknowledged that these exceptions must have their own limits, and preserved judicial review “of constitutional claims or questions of law.”²²⁸ It is thus left to courts to conclude, and for petitioners to argue, that what may appear to be a question of fact is really a question of law or a mixed question of fact and law. This may be enough flexibility to make the limitations of judicial review manageable in practice by allowing courts to avoid glaring miscarriages of justice. But it is not a fully satisfying explanation for why Congress has the authority to limit the judiciary’s jurisdiction over factual adjudication in the first place.

224. *Crowell*, 285 U.S. at 49–50.

225. Merrill, *supra* note 55, at 997–98.

226. Immigration and Nationality Act § 242(a), 8 U.S.C. § 1252(a) (2005).

227. *Id.* § 1252(a)(2)(b)–(c).

228. *Id.* § 1252(a)(2)(d).

V. THE INARTICULATE FACTORS

So far, we have seen that the deference doctrine appears in black-letter form to apply across the board, yet it is subject to considerable ambiguities that give courts leeway as to how to apply it in particular contexts. The substantial evidence rule applies in many different types of administrative appeals, but it may not mean the same thing in each type of case or even to two judges hearing the same case. While some of this judicial discretion is nontransparent, courts often discuss why particular types of agency decisions are owed more or less deference, even in applying what appears to be a uniform standard of review. In 1941, the Federal Committee on Administrative Procedures observed that courts are influenced by the “[t]he character of the administrative agency, . . . the nature and consequences of the administrative action, the confidence which the agency has won, . . . [and] the nature of the proceedings before the . . . agency.”²²⁹

In other words, courts deciding how much deference to give an agency’s adjudication pay attention to the agency’s expertise and to the weight of the interest at stake. The Committee described these as “inarticulate factors,”²³⁰ not so much because they are not spoken of openly, but because they are not included in formal articulations of the deference doctrine, even though there are clear logical justifications for their relevance. These inarticulate factors explain why immigration cases have been such a challenge for appellate courts because courts have lost some confidence in the competence of administrative immigration adjudication and the stakes in immigration cases are especially high.

A. Agency Expertise

The idea that executive agencies have expertise in the areas of their own mandates is, in effect, a version of the accuracy rationale because it suggests that agencies are more likely than courts to be able to make correct factual determinations within their fields of expertise. Recent writings by Justice Breyer provide a useful analytical approach for understanding the role of agency expertise in judicial deference, but they also open a revealing window on weaknesses of

229. FINAL REPORT OF THE ATTORNEY GENERAL’S COMMITTEE ON ADMINISTRATIVE PROCEDURE 91 (1941).

230. *Id.*

the prevailing doctrine. Writing in a private capacity, Justice Breyer has stressed “comparative expertise” as a decisive factor in understanding how courts relate to administrative agencies.²³¹ At the core of this idea is a conception of the tasks the judiciary performs well relative to other branches of government. Breyer writes:

Courts ask which institution, court, or agency is comparatively more likely to understand the critical matters that underlie a particular kind of legal question, broadly phrased. Courts are more likely to have experience with procedures, basic fairness to individuals, and interpreting the Constitution. Thus, where questions of this kind are at issue, courts are less likely to give much deference to agency decisions.²³²

This understanding of what courts do well dovetails closely with the classic fact-law distinction, and is hardly a new idea.²³³ But courts are just one side of the equation. The other critical question is, what do executive agencies do well? Breyer observes that “the public has less confidence in agency expertise” today than during the New Deal era.²³⁴ He notes that administrative policies are often set by “political appointees, not experts,”²³⁵ and that the resulting policy “often reflects political, not simply ‘scientific’ considerations.”²³⁶ He goes on to critique bureaucratic “tunnel vision,” overconfidence, and occasional lack of common sense.²³⁷ Breyer also acknowledges the diverse types of executive agencies, noting that “they come in different shapes and sizes” and that “[i]t is important to keep their size, complexity, and diversity in mind.”²³⁸

These observations point logically toward applying comparative expertise analysis in a fairly nuanced way, where a great deal would depend on the relative strengths of the particular agency at hand

231. See Breyer, *supra* note 20, at 2193–95.

232. *Id.* at 2193.

233. See Davis, *supra* note 14, at 418 (“For an accountant to exercise judgment on an accounting problem, or a mechanic on a mechanical problem, or an entomologist on a problem of policy under the Plant Quarantine Act, is like the exercise by a judge of his judgment on a question on which he is especially skilled, a question of law.”).

234. Breyer, *supra* note 20, at 2195.

235. *Id.*

236. *Id.* This observation undermines the premise that agencies deserve deference because they have technical expertise, but it may strengthen the assertion that executive agencies are in fact more politically accountable than often assumed. See Kagan, *supra* note 211, at 2331–32.

237. Breyer, *supra* note 20, at 2195.

238. *Id.* at 2191.

and on the type of question presented in an individual case. But it is difficult to square this fairly straightforward observation that not all agencies are created equal with Justice Breyer's broad assertion that "[a]gencies . . . are more likely to have experience with facts and policy matters related to their administrative missions. Thus courts will likely give agencies considerably more deference when decisions are about these matters."²³⁹ When writing for the majority of the Court, Breyer has said that "principles of judicial review . . . counsel judges to give expert agencies decision-making leeway in matters that invoke their expertise."²⁴⁰ It seems inconsistent to recognize that agencies are diverse and that some are more competent than others, and then to apply a sweeping rule of deference based on an assumption that all agencies are necessarily experts all of the time. But, as we have seen, this kind of one-size-fits-all rule is typical of the standard of review doctrine. It is left to lower courts to determine how to apply the across-the-board rule to the specifics of individual cases.

Nevertheless, Breyer's comparative expertise approach points toward a more structured approach. Let us assume three facts that Breyer highlights: (1) courts are good at ensuring a fair adjudication in a general sense, but are not technical specialists; (2) executive agencies are diverse; and (3) agencies sometimes have specialized expertise, but also sometimes display less admirable traits.²⁴¹ These premises would support varying the standard of judicial review in different types of cases according to two variables. First, is the question at hand one that calls simply for a generally fair hearing, or does it demand unique technical expertise? Second, is there reason to believe or to doubt that the agency in question actually has more useful expertise than a generalist court would have?²⁴²

There is good reason to think that courts, including the Supreme Court, apply this kind of comparative analysis implicitly. For example, the Supreme Court has said that judicial review of agency decisions should be "most deferential" when the subject is a scientific determination.²⁴³ Specialized agencies are most likely to have comparative advantages over courts in understanding the facts of an is-

239. *Id.* at 2193.

240. *Zadvydas v. Davis*, 533 U.S. 678, 700 (2001) (recognizing that the Executive Branch possesses "greater immigration-related expertise").

241. See Breyer, *supra* note 20, at 2195-96.

242. See, e.g., Anne R. Traum, *Constitutionalizing Immigration Law on Its Own Path*, 33 *CARDOZO L. REV.* 491, 522-30 (2011) (describing the immigration courts' lack of expertise in understanding the criminal records of immigrants subject to potential removal).

243. *Baltimore Gas & Elec. Co. v. Nat'l Res. Def. Council*, 462 U.S. 87, 103 (1983).

sue when highly technical matters of evolving science are at issue.²⁴⁴ Similarly, there is reason to think that courts are more or less deferential depending on their own comfort level with particular subject areas.²⁴⁵ According to a recent study by David Zaring, the Court of Appeals for the D.C. Circuit is less likely to affirm decisions by agencies that appear before it frequently.²⁴⁶ This is consistent with the hypothesis that, as judges gain more exposure to an area of law, they feel less need to defer to the relative expertise of the agency.²⁴⁷

The theory that all executive agencies have expertise worthy of deference has been severely tested in immigration adjudication, where courts of appeal have issued a number of decisions expressing concern about the capacity of the Executive Office for Immigration Review to adjudicate cases effectively.²⁴⁸ In a recent case, the Third Circuit acknowledged “the plight of immigration judges shoveling back a sea of cases,” with a footnote reference to reports about the under-resourcing of immigration courts.²⁴⁹ But it made this observation in a decision sharply rebuking an immigration judge for a “cavalier approach” to credibility assessment in an asylum case, remanding with a recommendation that the case be assigned to a different immigration judge and noting in another footnote that the immigration judge in question had “deeply troubled” the court in an earlier case as well.²⁵⁰

The thrust of this decision, certainly, is that the Third Circuit is not likely to defer much to decisions of this particular decision maker. However, what is the significance of the systemic strains on all immigration judges? It might be read simply as a note of sympathy and an attempt to soften the sting of the decision. But it also may be read as an explanation for why the court might not, in practice, grant any immigration judge’s decisions full deference because it

244. *But see* Emily Hammond Meazell, *Super Deference, the Science Obsession, and Judicial Review as Translation of Agency Science*, 109 MICH. L. REV. 733, 778–84 (2011).

245. Another flip side of expertise would be situations where an agency has self-interested reasons to apply the law in a particular way, which should lead courts to give less deference to its decisions. *See* Timothy K. Armstrong, *Chevron Deference and Agency Self-Interest*, 13 CORNELL J.L. & PUB. POL’Y 203, 206–34 (2004).

246. Zaring, *supra* note 17, at 183–84 (noting that agencies appearing before the D.C. Circuit fewer than ten times from 2000 to 2004 prevailed 80% of the time, compared to 68% for agencies appearing before that court more than ten times).

247. *Pierce*, *supra* note 21, at 88.

248. *See, e.g.*, *Abulashvili v. Attorney Gen.*, 663 F.3d 197, 208–09 (3d Cir. 2011); *Dia v. Ashcroft*, 353 F.3d 228, 250–51 (3d Cir. 2003).

249. *Abulashvili*, 663 F.3d at 208–09.

250. *Id.* at 205, 209.

may be reasonable to doubt their collective capacity to reach accurate decisions while laboring under so much strain. Consider this passage from another Third Circuit case, *Dia v. Ashcroft*, an asylum case where the claimant said he feared death or torture:

Repeatedly, we are left wondering how the IJ reached the conclusions she has drawn. Her opinion consists not of the normal drawing of intuitive inferences from a set of facts, but, rather, of a progression of flawed sound bites that gives the impression that she was looking for ways to find fault with Dia's testimony.

. . . [T]he inferences drawn and conclusions reached are in some instances non sequiturs, and in others, counterintuitive. The flow of the reasoning process appears to break down as the IJ, repeatedly, draws an unreasonable conclusion from a fact susceptible to differing interpretations. . . . [T]hey are an aggregation of empty rationales that devolve into an unsupported finding of adverse credibility. . . . Here, the conclusions of the IJ are more puzzling than plausible, more curious than commonsense.²⁵¹

These are just two decisions related to just two immigration judges. But they reflect deeper systemic doubts that have been articulated more directly by others, including the Attorney General.²⁵² Probably the most frequently cited systemic rebuke of the EOIR comes

251. 353 F.3d at 250-51. Among other things, Dia had testified that a group of men who were looking for him had come to his home and raped his wife, after which his wife begged him to flee. The IJ doubted the credibility of this account because Dia said he did not know why they had raped his wife, and doubted that his wife would have asked him to escape without her. The court devoted around two full pages to critiquing just this part of the IJ's reasoning, and there were many other examples. *Id.* at 254-55. *But see id.* at 266 (Alito, J., dissenting) ("[T]he IJ's belief about how a couple would likely react under such circumstances is just the sort of 'background knowledge' about human behavior that a fact finder is entitled to consider in evaluating a witness's credibility.").

252. See Memorandum from the Attorney General to Members of the Board of Immigration Appeals (Jan. 9, 2006), available at [www.justice.gov/av/reading room/ag-010906-boia.pdf](http://www.justice.gov/av/reading%20room/ag-010906-boia.pdf) ("I have watched with concern the reports of immigration judges who fail to treat aliens appearing before them with appropriate respect and consideration and who fail to produce the quality of work that I expect from employees of the Department of Justice. While I remain convinced that most immigration judges ably and professionally discharge their difficult duties, I believe there are some whose conduct can aptly be described as intemperate or even abusive and whose work must improve For the aliens who appear before them, our immigration judges are the face of American justice Not all aliens will be entitled to the relief they seek. But I insist that each case be reviewed proficiently and that each alien be treated with courtesy and respect."), cited in *Ming Shi Xue v. Bd. of Immigration Appeals*, 439 F.3d 111, 114 n.3 (2d Cir. 2006).

from Judge Posner, explaining why the courts of appeal were remanding so many immigration removal decisions:

[T]he adjudication of these cases at the administrative level has fallen below the minimum standards of legal justice. Whether this is due to resource constraints or to other circumstances beyond the Board's and the Immigration Court's control, we do not know. . . . All that is clear is that it cannot be in the interest of the immigration authorities, the taxpayer, the federal judiciary, or citizens concerned with the effective enforcement of the nation's immigration laws for removal orders to be routinely nullified by the courts²⁵³

Extensive statistical analysis of asylum cases in particular has found "amazing disparities in grant rates," suggesting that "the most important moment in an asylum case is the instant in which a clerk randomly assigns an application to a particular asylum officer or immigration judge."²⁵⁴

These types of doubts about a particular agency's adjudication apparatus could be more directly incorporated into the deference doctrine by applying a comparative expertise analysis as implied by Justice Breyer that would be tailored to particular agencies and types of cases. If an agency has a mechanism for ensuring the high quality of its decisions, then deference would be easier to justify.²⁵⁵ If there are reasons for doubting the agencies' competence, deference would be less justified. Also important here is the nature of the case at hand and the degree to which it demands technical specialization. The central point in this analysis is that expertise is specific, not to be assumed, and that, to a certain extent, agencies must earn deference by giving courts reasons to be confident in their decisions.

B. *The Interest at Stake*

In the Third Circuit's *en banc* decision in *Dia v. Ashcroft*, the majority debated the application of the substantial evidence standard with then-Judge Alito, whose dissenting opinion I discussed earlier in

253. *Benslimane v. Gonzales*, 430 F.3d 828, 830 (7th Cir. 2005) (citations omitted).

254. Jaya Ramji-Nogales et al., *Refugee Roulette: Disparities in Asylum Adjudication*, 60 STAN. L. REV. 295, 296 (2007).

255. For an example of what one possible quality control mechanism might look like, see Mariano-Florentino Cuéllar, *Auditing Executive Discretion*, 82 NOTRE DAME L. REV. 227, 252-74 (2006).

Part II, Section B. The majority agreed with Alito that administrative adjudicators deserved “some leeway.”²⁵⁶ But the court juxtaposed this concern against the weighty interests at issue in an asylum case to insist that “[t]he process of drawing inferences cannot be left to whim, but must withstand scrutiny.”²⁵⁷ Other courts have also noted the high stakes in ordering a remand to administrative adjudicators in the asylum context.²⁵⁸ But, while courts often note the stakes, they are also bound by established standards of review in which this factor plays no formal role. One of the most striking gaps in the deference doctrine is that it ignores the stakes involved in an individual case even as it asks appellate judges to let potential errors stand.²⁵⁹

This is a marked contrast to procedural due process, where the weight of the interest at stake is central to determining how extensive a procedure must be to be sufficient. Likewise, the presence of weighty constitutional issues will tend to mitigate against deferential review in large part because of the law-fact distinction.²⁶⁰ In a case concerning possible indefinite detention at the discretion of the Attorney General, the Supreme Court said, “[t]he Constitution may well preclude granting ‘an administrative body the unreviewable authority to make determinations implicating fundamental rights.’”²⁶¹ While this does not directly speak to the standard of review on appeal, it does impose an important limitation on the legislative branch’s power to delegate adjudication entirely to an executive agency. The Court gives less deference to findings of fact in First Amendment cases because “accurate enforcement of some rights is more important than accurate enforcement of others.”²⁶² But, when courts note the weight of the interest at stake in an appel-

256. *Dia v. Ashcroft*, 353 F.3d 228, 250 (3d Cir. 2003) (en banc).

257. *Id.*

258. See *Alcius v. Holder*, 374 F. App’x. 583, 590 (6th Cir. 2010); *Uanreroro v. Gonzales*, 443 F.3d 1197, 1208 (10th Cir. 2006); *Liu v. Dep’t of Justice*, 455 F.3d 106, 117 (2d Cir. 2006); *Wang v. Attorney Gen.*, 423 F.3d 260, 270–71 (3d Cir. 2005) (citing *Dia*, 353 F.3d at 250, on the importance of having a clear rationale for a decision given the high stakes of an asylum and removal case); *Shahandeh-Pey v. INS*, 831 F.2d 1384, 1389 (7th Cir. 1984).

259. See Kevin R. Johnson, *Responding to the “Litigation Explosion”: The Plain Meaning of Executive Branch Primacy over Immigration*, 71 N.C. L. REV. 413, 417 (1993) (“[T]he Court reviewed orders to deport persons to countries where they allegedly would face political persecution as no different in principle from any other agency decision.”).

260. Cf. Adam Hoffman, *Corralling Constitutional Fact: De Novo Fact Review in the Federal Appellate Courts*, 50 DUKE L.J. 1427, 1430–31 (2001) (noting when application of law to particular facts raises weighty constitutional issues, de novo review is warranted).

261. *Zadvydas v. Davis*, 533 U.S. 678, 692 (2001) (quoting *Superintendent v. Hill*, 472 U.S. 445, 450 (1985)).

262. *Cooper*, *supra* note 83, at 661.

late review of an administrative finding of fact, it may technically be *obiter dicta*.

The nature of immigration cases creates great tension around appellate deference because, on the one hand, the human stakes in a wrong decision are especially severe—in an asylum case, it means putting someone in danger of fundamental human rights violation²⁶³—while, at the same time, general scarcity of evidence makes factual adjudication especially difficult.²⁶⁴ The Second Circuit elaborated on the reasons why this matters:

Asylum petitions of aliens seeking refuge from alleged persecution are among the hardest cases faced by our courts. They are not games. And, despite their volume, these suits are not to be disposed of improvidently, or without the care and judicial attention-by immigration judges, in the first instance, and by federal judges, on appeal-to which all litigants are entitled. We should not forget, after all, what is at stake. *For each time we wrongly deny a meritorious asylum application, concluding that an immigrant's story is fabricated when, in fact, it is real, we risk condemning an individual to persecution.* Whether the danger is of religious discrimination, extrajudicial punishment, forced abortion or involuntary sterilization, physical torture or banishment, we must always remember the toll that is paid if and when we err.

. . . Under the circumstances, it is not surprising that the position of overburdened immigration judges and overworked courts has become a matter of wide concern. It is bound to be such when trivial mistakes can unwittingly lead to flawed decisions with grave consequences.²⁶⁵

In this passage, the Second Circuit connects several factors in explaining the challenges posed by immigration adjudication. Most important, the court connected the issues of accuracy to the weight of the interest at stake. The basic point here—that high stakes make sound decision-making more essential and thus justify more appellate scrutiny—hardly seems contestable. We provide more proce-

263. See *Mitondo v. Mukasey*, 523 F.3d 784, 788 (7th Cir. 2008) (“Asylum cases pose thorny challenges in evaluating testimony. Applicants regularly tell horrific stories that, if true, show past persecution and a risk of worse to come. But these stories rarely are susceptible to documentary proof . . .”).

264. See generally Kagan, *supra* note 69.

265. *Ming Shi Xue v. Bd. of Immigration Appeals*, 439 F.3d 111, 113–14 (2d Cir. 2006) (emphasis added).

dural protections in death penalty cases than in traffic court. Would it not be natural, then, for appellate courts to grant less deference in an asylum case than in a Social Security disability claim? Yet, the standard of review doctrine that we have does not incorporate an evaluation of the weight of the interest at stake in determining the level of appellate review. The black-letter rule is more mechanical. In administrative cases, findings of fact are reviewed for substantial evidence. Period.

VI. DEFERENCE AS A BALANCE

Immigration appeals incorporate several factors that highlight inherent limits of the rationales normally offered for appellate deference. The stakes in immigration cases are typically quite high. This means that on the one hand there is less reason to think the lower chance of success will deter would-be appellants; while, on the other hand, some judges are likely to be more resistant to affirming decisions about which they have doubts. At the same time, systemic weaknesses of the Board of Immigration Appeals give appellate judges less reason to assume the agency decisions they review are correct. These factors thus place considerable stress on the theory that observing demeanor will allow an immigration judge to correctly assess witness credibility. Congress has attempted to bolster this rule through the REAL ID Act. But, as we have seen, there is good reason to be concerned that reliance on demeanor will do little to increase the accuracy of adjudications.

It is thus not surprising that some judges in immigration appeals rely on the inherent ambiguities of the deference doctrine and appear not to defer to the BIA much at all. As we have seen, this ambiguity and the judicial flexibility that comes with it were built into the deference doctrine from early in the twentieth century. Within limits, judicial discretion can be useful because, even when stated rules are indeterminate, judicial decisions can still be substantially consistent.²⁶⁶ Judges can be guided by what Karl Llewellyn called "Situation-Sense," in which they digest facts and cases under a common pressure to produce "a satisfying working result."²⁶⁷ But a problem arises when this judicial flexibility runs counter to the stat-

266. See Dan M. Kahn, *Neutral Principles, Motivated Cognition, and Some Problems for Constitutional Law*, 125 HARV. L. REV. 1, 13-16, 77 (2011) (discussing multiple authors' responses to the indeterminacy or neutrality crisis).

267. KARL N. LLEWELLYN, *THE COMMON LAW TRADITION: DECIDING APPEALS* 60 (1960).

ed terms of formal rules. When the gap between formal rule and actual results grows, law becomes more vulnerable to the realist critique that political ideology and personal subjectivity are determining judicial outcomes. To escape this trap,²⁶⁸ it is important to consider actual practice, not only explicitly stated rules, in determining the content of law.²⁶⁹

My contention is that the ambiguity of the deference doctrine is a vivid example of judges finding situational and practical guides to steer them in the application of a legal rule. My goal is to better integrate the practice and the black-letter doctrine by better articulating the factors that many judges are implicitly using already to administer ambiguity. By doing this, the formal rules of law will be rendered more workable, less ambiguous, and more responsive to the real world, while judicial discretion and flexibility will be more clearly disciplined by a transparent analytical structure.

One might naturally conceive of appellate review as part and parcel of procedural due process since the degree to which an appellate body will scrutinize a decision is an essential characteristic of any adjudicatory procedure. As practiced by some courts, the deference doctrine has strong hints of modern procedural due process, especially when courts are influenced by the weight of the interest at stake to raise or lower the level of deference afforded an agency decision. Yet, by applying the same formal standard of review across a diverse spectrum of cases, the deference doctrine runs counter to the Supreme Court's warning against having a fixed concept of procedural due process for all types of adjudication.²⁷⁰ Standard of review doctrine has developed on an oddly separate track from procedural due process, in large part because deference in administrative law developed from a nineteenth-century legal era when civil procedure generally was more formalistic.²⁷¹ Today, American law has far more experience developing doctrines that achieve the flexibility necessary to adapt to different types of cases while still imposing

268. See Victoria Nourse & Gregory Shaffer, *Varieties of New Legal Realism: Can a New World Order Prompt a New Legal Theory?*, 95 CORNELL L. REV. 61, 123 (2009) ("One of the reasons that the old realism floundered . . . was the grossly exaggerated idea that law could be reduced to politics. Any new legal realism must anticipate this objection and respond to it. In our view, new legal realism must refuse to reduce law to politics or vice versa, and it must recognize the simultaneity of law and politics as institutional and participatory practices.").

269. Brian Leiter, *Beyond the Hart/Dworkin Debate: The Methodology Problem in Jurisprudence*, 48 AM. J. JURIS. 17, 24-25 (2003).

270. *Mathews v. Eldridge*, 424 U.S. 319, 334 (1976).

271. Merrill, *supra* note 55, at 987-88.

analytical discipline.²⁷²

Courts have on occasion described the importance of access to a judicial appeal from an agency in due process terms, but they have not included standards of appellate review within the rubric of the procedural due process doctrine. The Supreme Court has cited²⁷³ Justice Brandeis's dissent in *Crowell*, where he argued that, "[u]nder certain circumstances, the constitutional requirement of due process is a requirement of judicial process."²⁷⁴ But courts also draw an ambiguous distinction between the fundamental due process right to be heard and the right to a meaningful appeal by an administrative appellate body.²⁷⁵ At the same time, the lack of a constitutional right to an *administrative* appeal does not in any way limit the right to judicial appeal.²⁷⁶ The Supreme Court has said "[t]he APA requires meaningful review; and its enactment meant stricter judicial review of agency fact-finding than Congress believed some courts had previously conducted."²⁷⁷

For present purposes, it is sufficient to observe that there is a close connection between appellate standard of review and procedural due process.²⁷⁸ The three factors that *Mathews v. Eldridge* established for procedural due process—weight of the interest at stake, risk of error (i.e., accuracy), and efficiency²⁷⁹—all apply to determining the right level of appellate review. Deference doctrine can be employed with more analytical rigor by applying a balancing test very similar to *Mathews*. This would allow courts to look at the specific circumstances of particular cases, assess whether there is good reason to think the agency in question is well-positioned to make the most accurate decision, and weigh the potential advantages of efficiency against the risks of an errant decision. As we have seen, some courts already suggest this kind of analysis, but, because it is not a formal part of the doctrine, the analysis is less structured and consistent

272. Consider, for example, the development of varying levels of scrutiny in Fourteenth Amendment cases in constitutional law.

273. Merrill, *supra* note 55, 996–97.

274. *Crowell v. Benson*, 285 U.S. 22, 87 (1932) (Brandeis, J., dissenting).

275. *Dia v. Ashcroft*, 353 F.3d 228, 242 (3d Cir. 2003) (“The ‘meaningfulness’ requirement of *Mathews* pertains to ‘the opportunity to be heard’ and the ‘manner’ in which one is heard, not to a review by an administrative appellate body.” (quoting *Mathews*, 424 U.S. at 333)).

276. *Id.* (“[The] right to ‘meaningful review’ . . . has been confined to the context of review by federal courts, and not extended to review by an administrative appellate body.”).

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

278. Cf. Traum, *supra* note 242, at 533 (appellate judicial review as an application of *Mathews* due process safeguards).

279. *Mathews*, 424 U.S. at 335.

than it might be.

There are objections to balancing tests such as this one, as T. Alexander Aleinikoff pointed out in his seminal article on the subject in 1987.²⁸⁰ For example, balancing factors “pit individual against governmental interests”²⁸¹ in a manner that obscures collective interests in protecting individual rights.²⁸² It is useful to look for ways to see common interests between individuals and the state.²⁸³ But doing so excessively can simply obscure the real issue at stake, especially when a court is asked to protect the rights of an unpopular or marginalized person.²⁸⁴ A more persuasive concern is that, although balancing tests can structure a judicial inquiry by identifying the factors to be considered, they do not provide courts with any objective criteria by which to weigh factors against one another.²⁸⁵ This means that balancing tests cannot eliminate subjectivity entirely. As one commentator writes, “balancing tends toward ad hoc decision making.”²⁸⁶ This may be somewhat of an exaggeration since it is probably only in a close case with two genuinely weighty interests in tension that the structure of the balancing test will not effectively prescribe a result. Nevertheless, efforts to overcome the indeterminacy of balancing can feed a tendency to fall back on quasi-scientific cost-benefit analysis, translating every interest at stake to a quantifiable cost and, in the process, failing to give voice to fundamental values of justice.²⁸⁷

At a broader level, an important response to the limitations of

280. Aleinikoff, *supra* note 59, at 972–75.

281. *Id.* at 981.

282. *Id.*; Charles H. Koch, *A Community of Interest in the Due Process Calculus*, 37 HOUS. L. REV. 635, 643–44, 658–62 (2000).

283. Koch, *supra* note 282, at 664–70.

284. Consider, for example, Aleinikoff’s critique of *Hudson v. Palmer*, 468 U.S. 517 (1985), where the Court attempted to balance individual privacy of a prison inmate against government interest in prison security: “Society has a general interest in preventing unwarranted governmental intrusions As a collective body, we are in the cell with Palmer; the interests at stake are not his alone.” Aleinikoff, *supra* note 59, at 981. In a similar manner, when police object that civil libertarians make it harder to enforce the law, the civil libertarian can retort that the Bill of Rights is part of the law. This is clever rhetoric, but it is perhaps too clever to be persuasive because it favors redefining the terms over dealing head on with the substance. Let us assume that there is a genuine collective interest in individual liberty. There is also a genuine collective interest in security and fighting crime, and this can be in tension with individual rights. At some point a balance must be struck, and at the end of the day it must be struck by the legislature, the executive, or by a court.

285. *Id.* at 972–76.

286. Koch, *supra* note 282, at 637.

287. Aleinikoff, *supra* note 59, at 992–95.

balancing tests can be found in the field of comparative law, as developed in a recent article by Israeli scholars Moshe Cohen-Eliya and Iddo Porat about the global prevalence of proportionality analysis in modern constitutional jurisprudence.²⁸⁸ Proportionality is a term used widely outside the United States for balancing tests. It describes how a court should weigh an individual right against an asserted government interest, ending with a test that asks whether the government's interests are proportionate to the rights violation.²⁸⁹ Proportionality analysis is more open-ended than categorical tests, whereby, if a matter is found to be of a certain type, then the result follows automatically. Cohen-Eliya and Porat's account of proportionality fully acknowledges that balancing tests do not predetermine results, which they argue is part of their appeal: "The flexibility inherent in proportionality provides judges with leeway to develop doctrine freely, and to decide when to intervene and when not, taking into account a wide range of considerations, including public opinion, the potential for political backlash, and institutional memory."²⁹⁰

What is odd about this is that this desire for flexibility echoes the appeal that American judges in the deference doctrine have advocated, even though the deference doctrine is highly categorical. If a finding is a question of fact, then the standard of review is deferential. But such categorical rules can also be indeterminate in practice because they can be ambiguous. By contrast, balancing tests are "standard-based;"²⁹¹ they channel analysis in useful ways, acknowl-

288. See generally Cohen-Eliya & Porat, *supra* note 223.

289. Cohen-Eliya and Porat state that proportionality is not used in American constitutional law. *Id.* at 465. I would suggest that this is true only in terms of terminology, with American legal culture more accustomed to thinking in terms of balancing tests. While the term proportionality is not embraced by American constitutional law, the essential substance of the analytical method would be immediately familiar to an American lawyer who used Fourteenth Amendment analysis. In proportionality analysis:

After the government has shown that its action, which infringed the constitutional right in question, had a legitimate purpose, courts across the globe undertake a proportionality analysis, which usually proceeds in three stages. First, it will examine whether the means that were applied further the legitimate governmental end (the rationality test); second, whether the government chose the least restrictive means to further that end (the necessity test); and third, whether the benefits of the governmental objective are proportionate to the violation of the constitutional right (the balancing test).

Id. at 464. This is exactly what the U.S. Supreme Court does when it asks whether a government action is narrowly tailored to a compelling state interest, sufficient to justify infringing a fundamental right.

290. *Id.* at 468.

291. *Id.*

edging the legitimacy of two competing concerns rather than categorically declaring one side to be right.²⁹² The advantage of a balancing test is that it copes with indeterminacy in a more straightforward manner and requires judges to conduct their reasoning in a prescribed manner as they work toward a conclusion.

The way in which balancing tests channel analysis leads to the most important insight offered by Cohen-Eliya and Porat. They argue that the twentieth-century shift toward proportionality analysis grew from changing conceptions of democratic government, from a “culture of authority” to a “culture of justification.”²⁹³ In the culture of authority, “[t]he legitimacy and legality of governmental action is derived from the fact that the actor is authorized to act.”²⁹⁴ This is what American courts focus on when they decide administrative law cases regarding delegation of powers, dwelling on the question of whether an agency has acted within its statutory mandate. By contrast, in a culture of justification:

[T]he existence of authorization to act is a necessary but not sufficient condition for legitimacy and legality. Rather, the crucial component in the legitimacy and legality of governmental action is that it is justified in terms of its “cogency” and its capacity for “persuasion,” that is, in terms of its rationality and reasonableness.²⁹⁵

This focus on whether an authority can give persuasive reasons for its decision making is an essential feature of American administrative law.²⁹⁶ But perhaps what is true for agencies should also be true for courts. If judges are to defer to others to make a high-stakes decision, persuasive reasons should justify the deference.

I do not suggest that, by having all courts balance factors, inconsistency will disappear and judicial judgment calls will play no role. But there would be a significant relative gain in the transparency of judicial reasoning as well as in the predictability of results. Deference makes sense, but only sometimes; it makes much more sense when the stakes are low, when the agencies are adequately resourced and more likely to be specially competent, and when efficiency is paramount. But, by the converse, when a great deal is at risk, when the agency’s track record for reliability is in question, and

292. *Id.* at 469–70.

293. *Id.* at 474–75.

294. *Id.* at 475.

295. *Id.*

296. *Id.* at 487–88.

when the administrative adjudicators are over-stretched, deference is a recipe for inaccuracy and tragedy.