



ARTICLE

Due Process and Mass Adjudication: Crisis and Reform

David Ames, Cassandra Handan-Nader,
Daniel E. Ho & David Marcus*

Abstract. *Goldberg v. Kelly* and its progeny imposed a judicial model for decisionmaking on much of the administrative state. The linchpin of procedural due process was accuracy: *Goldberg's* premise was that agencies could improve the accuracy of their decisionmaking by giving individuals the sort of procedural rights enjoyed in court. In the wake of the due process revolution, federal agencies now adjudicate more cases than all Article III courts combined, and state adjudicators handle millions of cases with court-like procedures in their administrative systems. Yet despite *Goldberg's* premise, mass adjudication has struggled to achieve an adequate threshold of accuracy. In much of the administrative state, this struggle has deepened into an urgent crisis. The leading academic response argues for a turn to "internal administrative law" and management techniques, not external law, to improve the quality of agency adjudication. Many agencies in turn have

* David Ames is Director, Appellate Operations, Bergmann and Moore, LLC; and Former Chief of Office of Quality Review, Board of Veterans' Appeals, U.S. Department of Veterans Affairs. Cassandra Handan-Nader is a Ph.D. student in Political Science; and Graduate Student Fellow, Regulation, Evaluation, and Governance Lab, Stanford University. Daniel E. Ho is the William Benjamin Scott and Luna M. Scott Professor of Law, Professor of Political Science, Senior Fellow at Stanford Institute for Economic Policy Research, and Faculty Director of the Regulation, Evaluation, and Governance Lab, Stanford University. David Marcus is a Professor of Law, UCLA School of Law.

We thank Anne McDonough, Oluchi Mbonu, and Reid Whitaker for terrific research assistance; our many anonymous interview subjects across agencies for insightful conversations; Michael Asimow, Dan Bussel, Andy Coan, Blake Emerson, Jacob Goldin, David Hausman, Mark Krass, Daryl Levinson, Rob MacCoun, Jon Michaels, Anne Joseph O'Connell, Nick Parrillo, Richard Re, James Ridgway, Sam Sherman, Bill Simon, Mila Sohoni, Angela Teuscher, Zac Townsend, and Eugene Volokh; and participants at the Munro Distinguished Lecture in Stanford's political science department, the faculty fellow workshop at the Center for Advanced Study in the Behavioral Sciences, the faculty workshop at Northwestern Pritzker School of Law, the faculty workshop at the University of San Diego School of Law, the faculty workshop at SMU Dedman School of Law, the faculty workshop at Stanford Law School, the faculty workshop at UCLA School of Law, and the C. Boyden Gray Center's conference on "The Veterans Appeals Process: A Case of Administrative Crisis and Possible Reforms" for helpful comments and conversations.

Due Process and Mass Adjudication
72 STAN. L. REV. 1 (2020)

responded with such quality assurance programs, but we know next to nothing about how such programs have evolved, how they function, and whether they work.

Our Article is the first to rigorously investigate the promise and pitfalls of quality assurance as a guarantor of accuracy in agency adjudication. We make three contributions. First, we use in-depth interviews with senior agency officials and a wide array of internal agency materials to document the evolving use of quality assurance at three federal agencies whose mass adjudication epitomizes *Goldberg's* domain. This history documents years of fits and starts, as agencies tried to manage what is commonly referred to as a "quantity-quality" tradeoff. It also reveals deep tensions and ambiguities in what the agencies intend as the purpose of quality assurance.

Second, we provide the first rigorous test of quality assurance, the leading academic response to *Goldberg's* limitations. We use a rich dataset, never before available to outside academics, of over 500,000 cases decided by the Board of Veterans' Appeals (BVA) to craft a rigorous evaluation of a natural experiment created by its "Quality Review" program. Under this program, cases were randomly selected for review of draft decisions by an elite squadron of attorneys to correct substantive legal errors. BVA used this program ostensibly to reduce appeals to and remands from the courts reviewing its decisions. We show that the program failed on its own terms: Cases selected for Quality Review fared no better than cases that were not. BVA used the program not to vindicate *Goldberg's* premise, but to mollify external oversight bodies, most notably Congress, with the appearance of accuracy.

Third, our historical and empirical evidence has substantial implications for major theoretical debates about "internal administrative law" and the emerging crisis in mass adjudication. We show that conventional scholarly accounts are in need of much refinement. Deficiencies in mass adjudication will not be fixed solely through external constitutional law, with courts imposing remedies from outside. Nor will they be fixed solely by internal administrative law. *Goldberg's* original premise of decisional accuracy requires a hybrid of external intervention, stakeholder oversight, and internal agency management. We offer concrete policy prescriptions, based on a pilot one of us designed as BVA's Chief of the Office for Quality Review, for how quality assurance might be re-envisioned to solve the looming crisis of decisional quality.

Table of Contents

Introduction	4
I. The Looming Crisis of Decisional Quality	9
A. Three Pivotal Agencies	9
1. Social Security disability benefits adjudication	12
2. Immigration adjudication	13
3. Veterans' benefits adjudication	14
B. The Caseload Crisis and the Quantity-Quality Tradeoff	16
II. Systemic Quality, Constitutional Obligation, and Internal Administrative Law	20
A. The Limits of External Law	20
1. Accuracy, constitutional obligation, and remedial mismatch	20
2. Jurisdiction channeling	25
B. Internal Administrative Law and Quality Assurance	28
III. Quality Assurance in Historical Perspective	30
A. SSA's Decades of Experiments	31
1. Initial efforts	31
2. Bellmon Review	33
3. SSA's modern quality assurance system	36
B. EOIR's Inattention	40
C. BVA and Quality Review	42
IV. Empirical Case Study: BVA's Quality Review Program	48
A. A Critical Test Case	48
B. Results and Analysis	50
V. Implications	56
A. Theoretical Implications	57
1. Internal administrative law and the threat from outside	57
2. Internal pressures and distortions	59
a. Resistance problems	60
b. Career and agency incentives	61
c. Conflicting goals	64
B. Improving Quality Assurance: Externally Influenced Internal Administrative Law	67
1. Mandating disclosure of methodology	69
2. Institutionalizing oversight	70
3. Improving design of quality assurance	71
4. Structural reform litigation	73
Conclusion	77

Introduction

Federal administrative adjudication is in distress. When a veteran applies for benefits and loses, more than five years will typically pass before his appeal gets decided.¹ Judges handling these appeals must decide twenty-five to thirty cases per week lest the backlog worsen.² This pace is crushing. A judge often has no more than an hour to review thousands of pages in the record.³ In 2017, about 100 agency attorneys issued a public “loss of confidence” statement, declaring that increased caseloads prevented them from reviewing appeals as closely as applicable law requires (de novo review).⁴ “I could have integrity here or I could stay employed,” one judge confided, describing the pressure to issue many decisions and the impact it has on the quality of his decisionmaking.⁵

Immigration judges likewise labor under immense backlogs, with caseloads quadrupling from 2001 to 2017.⁶ The quality of decisions prompted scathing critique, including then-Judge Richard Posner’s takedown of “the Immigration Court” as “the least competent federal agency,”⁷ and the American Bar Association’s condemnation of immigration adjudication as “irredeemably dysfunctional and on the brink of collapse.”⁸ The Social Security disability benefits adjudication system similarly struggles, with long waits for hearings⁹

-
1. 2017 BOARD VETERANS’ APPEALS ANN. REP. 25.
 2. James D. Ridgway, Opening Remarks at the C. Boyden Gray Center for the Study of the Administrative State Conference: The Veterans Appeals Process; A Case of Administrative Crisis and Possible Reforms 15 (Oct. 17, 2018) (presentation on file with authors).
 3. *See id.* at 13, 15 (noting the number of electronic documents per appeal and the review time for each appeal).
 4. Letter from Douglas E. Massey, Counsel, Bd. of Veterans’ Appeals, to David Shulkin, Sec’y of Veterans Affairs 1 (Sept. 18, 2017), <https://perma.cc/K469-2MPG>.
 5. Telephone Interview with Interview Subject No. 4 (Sept. 11, 2018). We interviewed current and former Board of Veterans’ Appeals personnel and others with inside knowledge of the veterans’ benefits adjudication system. We promised interview subjects anonymity, and thus throughout this Article we use generic identifiers when citing to interview transcripts.
 6. Nat’l Ass’n of Immigration Judges, NAIJ Has Grave Concerns Regarding Implementation of Quotas on Immigration Judge Performance Reviews 3 (2017), <https://perma.cc/7HWC-7AKK>.
 7. *Chavarria-Reyes v. Lynch*, 845 F.3d 275, 280 (7th Cir. 2016) (Posner, J., dissenting).
 8. 2 COMM’N ON IMMIGRATION, AM. BAR ASS’N, 2019 UPDATE REPORT: REFORMING THE IMMIGRATION SYSTEM 2-3 (2019), <https://perma.cc/8FRK-ARX3>.
 9. For wait times by hearing office, see *Average Wait Time Until Hearing Held Report (for the Month of August 2019)*, SOC. SECURITY ADMIN., <https://perma.cc/TUV2-SGL6> (archived Oct. 16, 2019).

and rushed, poor decisionmaking excoriated by federal judges.¹⁰ In 2018, one federal judge lamented that “the Social Security system is broken.”¹¹

An onslaught of cases threatens the capacity of these agencies—the Social Security Administration (SSA), the Department of Veterans Affairs (VA), and the Department of Justice—to render accurate decisions. This crisis of decisional quality has major policy implications for the rights of immigrants to asylum, the rights of veterans to just compensation for their service, and disabled workers’ access to the social safety net. But its significance extends across the administrative state. *Goldberg v. Kelly* demands procedures modeled on court-based litigation for a myriad of areas involving mass administrative decisionmaking.¹² For any of these systems, whether they involve welfare benefits, access to Medicaid, or applications for patent rights—that is, the vast terrain of quasi-adjudication in the administrative state—a crush of caseloads can undermine access to accurate decisionmaking.

The crisis of decisional quality in administrative adjudication seems to beg for a constitutional response. Under *Mathews v. Eldridge*’s famed approach to due process, an individual’s entitlement to enhanced process ostensibly increases as the risk of error rises.¹³ But the Supreme Court’s willingness to recognize new procedural rights for individuals caught in the Bleak House of agency adjudication quickly dissipated after *Mathews*.¹⁴ Even if the Due Process Revolution of the 1970s had not long since crested,¹⁵ there would be reason to doubt the remedial efficacy of enhanced procedural rights, as we spell out below.¹⁶

10. See, e.g., *Wallace v. Colvin*, 193 F. Supp. 3d 939, 941 (N.D. Ill. 2016); *Batista v. Colvin*, No. 13 Civ. 4185, 2014 WL 2618534, at *2-3 (E.D.N.Y. June 12, 2014); *Freismuth v. Astrue*, 920 F. Supp. 2d 943, 954 (E.D. Wis. 2013).

11. *Williams v. Berryhill*, No. 0:17-1203, 2018 WL 3569076, at *2 (D.S.C. July 25, 2018).

12. 397 U.S. 254, 266-71 (1970). It is well-established that while the *Goldberg* Court noted that procedures “need not take the form of a judicial or quasi-judicial trial,” *id.* at 266, the due process explosion has judicialized forms of agency adjudication, see Henry J. Friendly, “Some Kind of Hearing,” 123 U. PENN. L. REV. 1267, 1269 (1975) (“[T]he tendency to judicialize administrative procedures has grown apace in the United States. English judges and scholars consider that we have simply gone mad in this respect.” (footnote omitted)).

13. See 424 U.S. 319, 335 (1976).

14. See Thomas W. Merrill, *Jerry L. Mashaw, the Due Process Revolution, and the Limits of Judicial Power*, in ADMINISTRATIVE LAW FROM THE INSIDE OUT: ESSAYS ON THEMES IN THE WORK OF JERRY L. MASHAW 39, 43-45 (Nicholas R. Parrillo ed., 2017); Richard J. Pierce, Jr., Essay, *The Due Process Counterrevolution of the 1990s?*, 96 COLUM. L. REV. 1973, 1981-84 (1996).

15. See Pierce, *supra* note 14, at 1989.

16. See *infra* Part II.A.1.

Another remedy may hold more promise. In 1974, the luminary Yale professor Jerry Mashaw famously criticized *Goldberg* and the Due Process Revolution precisely for their incapacity to ensure systemic accuracy in agency adjudication.¹⁷ He suggested an alternative to individual procedural rights, arguing that agencies ought to have an obligation to develop standards for decisional quality, to evaluate decisionmaking against these standards, and to develop responsive managerial interventions to improve adjudicator performance agencywide.¹⁸ Mashaw made a “quality assurance system” a key plank in the “internal law of administration” he described and defended in *Bureaucratic Justice*,¹⁹ a book that remains the canonical account of justice in agency adjudication.

In recent years, “internal administrative law”—the corpus of rules, evaluative strategies, and management techniques agencies use to govern themselves—has risen to the top of many scholars’ research agendas.²⁰ Despite this attention, and despite the dire policy need that crises of decisional quality pose, administrative law scholarship has ignored quality assurance initiatives for mass decisionmaking by agency adjudicators. Our Article fills this void. It is the first academic study in nearly forty years of the quality assurance initiatives that agencies have used to evaluate and improve decisional quality for systems of mass adjudication.²¹

We make three chief contributions. First, we use material generated by Freedom of Information Act (FOIA) requests, in-depth interviews with senior agency officials, historical records, and other sources to provide a rich history and account of the various initiatives that the major systems of federal administrative adjudication have used to evaluate and try to improve their judges’ decisional quality. Promoted by institutions like the Administrative Conference of the United States and the U.S. Government Accountability

17. See Jerry L. Mashaw, *The Management Side of Due Process: Some Theoretical and Litigation Notes on the Assurance of Accuracy, Fairness, and Timeliness in the Adjudication of Social Welfare Claims*, 59 CORNELL L. REV. 772, 823 (1974).

18. See *id.* at 791.

19. JERRY L. MASHAW, *BUREAUCRATIC JUSTICE: MANAGING SOCIAL SECURITY DISABILITY CLAIMS* 15, 149 (1983).

20. E.g., Christopher J. Walker, *Administrative Law Without Courts*, 65 UCLA L. REV. 1620, 1624 (2018) (“Internal administrative law . . . has become a hot subfield in administrative law.”).

21. See MASHAW, *supra* note 19, at 145-68 (describing SSA’s control and oversight of the disability benefits adjudication process). Various government entities, including the now-Government Accountability Office (formerly the General Accounting Office) and agency inspectors general, have investigated and critiqued quality assurance programs. We cite their reports and analyses throughout this Article. But we are unaware of academic studies that have followed up on Mashaw’s work from the 1970s and 1980s.

Office (GAO), these initiatives are “internal administrative law” in that they can exist distinct from external legal oversight.

Second, we draw upon a rich internal database of over 500,000 decisions by VA judges to evaluate a prototypical initiative of internal administrative law. The Board of Veterans’ Appeals (BVA) operated its “Quality Review Program” without formal change for fifteen years. It randomly selected cases for review, creating a unique natural experiment. The dataset, which the agency itself uses to manage its adjudicatory process, has never before been publicly studied. We use this dataset to examine whether the program, which an elite detail of four to six full-time attorneys implemented for over fifteen years, improved decisional quality.²² Through in-depth investigation, we show that the program generated an all-but-meaningless measure of decisional quality. The program failed to identify errors in decisionmaking in any rigorous way, but it did generate an “accuracy rate” the agency used to defend its work.²³ Quality Review (QR), in essence, became public relations. Our findings reveal the extent of the crisis threatening the federal administrative state: Even an agency that claims success may in fact be hiding dramatic declines in decisional quality.

Third, we derive broad theoretical and policy implications from these findings. The important example of quality assurance challenges leading accounts of how internal administrative law can best develop and function. These accounts argue that agencies should be given more room to govern themselves than reviewing courts tend to allow, faulting judicial review for its disruptive effect on internal administrative law. We show that these accounts gloss over pressures that can distort internal administrative law’s evolution, pressures that BVA’s QR Program and other experiments with quality assurance highlight. We argue that a more nuanced relationship between external and internal sources of administrative governance can counteract these pressures. This relationship would allow internal administrative law to develop in ameliorative ways but keep it from drifting in pathological directions. We derive concrete policy prescriptions for how internal and external sources of agency governance can productively interact. We explain how legislators and courts can prompt agencies to design and administer successful quality assurance initiatives. Even agencies laboring under huge caseloads can take steps to protect decisional quality and thereby meet their constitutional obligations.

22. For a description of the program, see notes 302-14 and accompanying text below.

23. BVA variously uses the terms “accuracy rate” and “accuracy rating.” *E.g.*, 2017 BOARD VETERANS’ APPEALS, *supra* note 1, at 15 (using “accuracy rate”); 2016 BOARD VETERANS’ APPEALS ANN. REP. 7 (using both “accuracy rating” and “accuracy rate”). We use the term “accuracy rate.”

Our findings also challenge conventional wisdom in administrative law. First, formalism and functionalism matter for the standard of review. While BVA's standard of review remained formally the same, QR functionally morphed into a rubber stamp in light of the goals of reporting high accuracy rates. Second, institutional design matters for the standard of review. The fact that attorneys—not judges—were conducting QR likely accounts for much of its functional weakness. The specter of staff attorneys' removal, coupled with the lure of appointment and promotion, may influence the stringency of review as much as a standard's formal label. Third, while reviewing courts can correct erroneous decisions, appellate review appears ineffective at systematic error correction. By comparing errors flagged by the QR team and cases selected for further appeal, it does not appear that claimants are able to effectively select disputes likely to contain reversible error. Last, if we are to take internal administrative law seriously, our findings illustrate the importance of taking research beyond formal reports and guidance. The historical experience, particularly at SSA, reveals deep tensions and tradeoffs around the design of such internal governance structures. And as our study illustrates, scholars must get inside agencies to understand the critical ways in which internal law actually operates.

Our Article proceeds as follows. In Part I, we introduce three agencies that are, in many ways, the most prominent examples of mass adjudication in the federal administrative state: VA's Board of Veterans' Appeals, SSA's Office of Hearings Operations, and the Department of Justice's Executive Office for Immigration Review. Part II explains that, while the caseload crises faced by these agencies may be matters of constitutional significance, traditional constitutional and administrative law do not offer promising remedies. The leading alternatives are quality assurance initiatives—the internal training, guidance, monitoring, and continuous improvement programs agencies use to manage and improve accuracy.²⁴ In Part III, we survey four mostly unexamined decades of agency experiences with such programs. We show that the programs occur in fits and starts, are heavily contested, and expose deep ambiguities as to the purposes of adjudication. In Part IV, we provide the results of our rich empirical study that leverages random case selection to examine the impact of BVA's QR program. We show that on its own terms, the program failed and was co-opted to generate the *appearance* of effectiveness. Part V identifies the theoretical and policy implications of our findings for internal administrative law and for agencies' efforts to meet an adequate threshold of decisional quality.

24. See Daniel E. Ho & Sam Sherman, *Managing Street-Level Arbitrariness: The Evidence Base for Public Sector Quality Improvement*, 13 ANN. REV. L. & SOC. SCI. 251, 252-53 (2017).

I. The Looming Crisis of Decisional Quality

Agencies handling large volumes of cases or claims have long weathered crises that have tested their capacities to render accurate decisions. In the early 1980s, for example, changes in disability policy and a vigorous effort to terminate benefits for those no longer disabled caused disability benefits dockets to explode and prompted a significant backlash against SSA in the federal courts.²⁵ The struggles adjudicators face to give welfare beneficiaries meaningful “fair hearings” are well documented.²⁶

But the present moment may be unprecedented. Caseload crises simultaneously challenge the capacity of SSA, BVA, and the Executive Office of Immigration Review (EOIR) to decide cases adequately. No single reason explains why otherwise unrelated dockets have expanded so rapidly at approximately the same time. The three agencies, which we introduce here, have responded in similar ways, including by asking their judges to increase the pace of decisionmaking. The quotas and case completion goals adjudicators face pose a systemic threat to decisional quality.

A. Three Pivotal Agencies

Hundreds of agencies across the United States adjudicate millions of cases, claims, applications, requests, petitions, and the like each year. We focus here on three of the most prominent sets of adjudicators in the federal administrative state: BVA’s veterans law judges (VLJs), SSA’s administrative law judges (ALJs), and EOIR’s immigration judges (IJs). Often bundled together for the analysis of administrative adjudication and its difficulties,²⁷ these three agencies usefully represent the challenges that *Goldberg*-style mass adjudication faces, for several reasons. First, these systems are unquestionably three of the most important adjudicatory agencies in the federal government. Together, their judges adjudicate more cases annually than do the nation’s federal district court and court of appeals judges.²⁸ SSA alone employs the vast majority of

25. See MARTHA DERTHICK, *AGENCY UNDER STRESS: THE SOCIAL SECURITY ADMINISTRATION IN AMERICAN GOVERNMENT* 5 (1990).

26. See, e.g., VICKI LENS, *POOR JUSTICE: HOW THE POOR FARE IN THE COURTS* 9-39 (2016) (describing welfare “fair hearings”).

27. See, e.g., Peter M. Shane, *Boundary Disputes: Jerry L. Mashaw’s Anti-Formalism, Constitutional Interpretation and the Unitary Presidency*, in *ADMINISTRATIVE LAW FROM THE INSIDE OUT*, *supra* note 14, at 188, 205; Paul Verkuil, *Meeting the Mashaw Test for Consistency in Administrative Decision-Making*, in *ADMINISTRATIVE LAW FROM THE INSIDE OUT*, *supra* note 14, at 239, 244-45.

28. During Fiscal Year (FY) 2017, the U.S. Courts of Appeals and the U.S. District Courts terminated 54,347 and 365,238 cases, respectively. *Judicial Caseload Indicators—Judicial Business 2017*, U.S. CTS., <https://perma.cc/CA7N-MT9C> (archived Oct. 17, 2019). In the same year, BVA decided 52,661 appeals. 2017 BOARD VETERANS’ APPEALS, *supra* note 1, at *footnote continued on next page*

federal ALJs.²⁹ Second, VLJs, IJs, and ALJs enjoy a quasi-judicial status that poses particular difficulties for quality assurance.³⁰ Some agency adjudicators are civil servants subject to control by bureaucratic superiors. Supervisors can tell them how to decide matters, and the agency can discipline them based on the quality of their work.³¹ ALJs, by contrast, enjoy guarantees of independence that strictly limit quality-based evaluation, performance management, and discipline.³² Although VLJs and IJs have less robust protections, their judicial roles insulate them from strict hierarchical control, and custom (if not formal regulation) limits decisional quality as a factor in performance evaluation.³³ *Goldberg* requires a degree of judicial independence for adjudication.³⁴ But it can be a double-edged sword for quality assurance and significantly complicate management.³⁵

Third, VLJs, IJs, and ALJs have similar tasks that justify an interagency comparison of quality assurance initiatives.³⁶ They often review decisions

30. IJs rendered 149,581 “initial case completions” in FY 2017. 2017 EXECUTIVE OFF. FOR IMMIGR. REV. STAT. Y.B. 10. And SSA ALJs completed 685,657 hearings that year. 2017-2019 SOC. SECURITY ADMIN ANN. PERFORMANCE REP. 35.

29. JACK M. BEERMANN & JENNIFER L. MASCOTT, RESEARCH REPORT ON FEDERAL AGENCY ALJ HIRING AFTER *LUCIA* AND EXECUTIVE ORDER 13843, at 2 (2019), <https://perma.cc/A283-VG93>.

30. As our discussion of SSA’s quality assurance efforts over several decades shows, ALJs have pushed back against programs that singled particular judges out for scrutiny, on grounds that doing so would interfere with decisional independence. *See infra* Part III.A.

31. *See, e.g.*, OFFICE OF THE INSPECTOR GEN., U.S. DEP’T OF COMMERCE, NO. OIG-15-026-A, U.S. PATENT AND TRADEMARK OFFICE: USPTO NEEDS TO STRENGTHEN PATENT QUALITY ASSURANCE PRACTICES 5 (2015), <https://perma.cc/LU6P-T29A> (describing quality as a factor in performance evaluation for patent examiners); *id.* at 8 (describing quality as a factor in examiner discipline).

32. *See* Kent Barnett, *Against Administrative Judges*, 49 U.C. DAVIS L. REV. 1643, 1655-56 (2016).

33. Results generated by BVA’s QR program, for instance, do not factor into VLJ performance evaluations. Telephone Interview with Interview Subject No. 3 (Aug. 31, 2018) (remarking that “it always bugged” the interviewee that QR “didn’t seem to factor into” a VLJ’s performance evaluation). Before 2007, IJs were exempt from performance evaluation. Nat’l Ass’n of Immigration Judges, *supra* note 6, at 3. Presently, IJs may be subject to “performance appraisals” based, in part, on “poor decisional quality,” but these appraisals must “fully respect[] their roles as adjudicators.” 8 C.F.R. § 1003.0(b)(v) (2019).

34. *See* *Goldberg v. Kelly*, 397 U.S. 254, 271 (1970).

35. *See infra* Part III. For a related account of the complexities associated with managing street-level bureaucracy, see generally Ho & Sherman, *supra* note 24.

36. For an example of a prominent scholar analyzing SSA, BVA, and EOIR adjudication together as cognate systems presenting similar management problems, see Verkuil, *supra* note 27, at 244.

Due Process and Mass Adjudication
72 STAN. L. REV. 1 (2020)

made in nonjudicial settings by non-attorney adjudicators;³⁷ they receive written evidence, develop cases in an inquisitorial manner, and can hold hearings;³⁸ they decide cases subject to review by both specialized and generalist appellate tribunals;³⁹ and each works for an agency that administers a single national policy through the one-by-one adjudication of huge numbers of cases.⁴⁰ To be sure, other agencies administer policy through ALJs or their equivalents.⁴¹ For some, modest caseloads make the obligation to meet an adequate threshold of decisional quality an entirely different challenge.⁴² Others have large dockets, but they handle cases with much lower stakes than cases that BVA, EOIR, and SSA adjudicate.⁴³

Finally, due process protections work similarly for veterans, immigrants, and workers. Each has a protected interest at stake when the agency adjudicates his or her claim.⁴⁴ Although the weight of the interest may differ from one context to the next,⁴⁵ the federal courts have described each type of interest in terms that recognize its fundamental importance to the three

37. See *infra* Part I.A.

38. *Id.*

39. *Id.*

40. Cf. Jerry L. Mashaw, *Organizing Adjudication: Reflections on the Prospect for Artisans in the Age of Robots*, 39 UCLA L. REV. 1055, 1059 (1992) (describing this challenge for SSA). For each system's caseload, see Part I.B below.

41. For more on ALJs and "administrative judges" generally, see Barnett, *supra* note 32, at 1652-62.

42. The five ALJs working for the Securities and Exchange Commission (SEC) rendered 115 "Initial Decisions" in 2018. *ALJ Initial Decisions: Administrative Law Judges*, U.S. SEC. & EXCHANGE COMMISSION, <https://perma.cc/277N-4TRD> (last updated Oct. 7, 2019); see also *Lucia v. SEC*, 138 S. Ct. 2044, 2049 (2018) (noting that the SEC employs five ALJs). The small number of ALJs and the comparatively modest per capita caseload means that the SEC faces quality challenges quite different from ensuring consistency across hundreds of decisionmakers, and from ensuring accuracy in decisions rendered under remarkable time constraints.

43. ALJs at the Office of Medicare Hearings and Appeals (OMHA) decide large numbers of cases, but the value of each case can be modest. See Jonah B. Gelbach & David Marcus, *Rethinking Judicial Review of High Volume Agency Adjudication*, 96 TEX. L. REV. 1097, 1104-05, 1104 tbl.1, 1105 n.33 (2018) (noting a \$160 amount-in-controversy threshold for OMHA ALJ hearings).

44. See, e.g., *Flores-Chavez v. Ashcroft*, 362 F.3d 1150, 1160 (9th Cir. 2004) (describing immigrants' liberty interests at stake in removal proceedings); *Flatford v. Chater*, 93 F.3d 1296, 1304-05 (6th Cir. 1996) (recognizing disability claimants' property interest in potential benefits); *Nat'l Ass'n of Radiation Survivors v. Derwinski*, 994 F.2d 583, 586, 588 n.7 (9th Cir. 1992) (recognizing veterans' property interest in benefits).

45. The Federal Circuit suggests, for instance, that an immigrant's right not to be deported weighs more heavily than a veteran's right to benefits. *Pitts v. Shinseki*, 700 F.3d 1279, 1285-86 (Fed. Cir. 2012).

categories of beneficiaries.⁴⁶ Thus, the consequences of an error—the consignment of a veteran, immigrant, or disabled worker to a laborious appeals process—are comparable across the three agencies.

1. Social Security disability benefits adjudication

SSA administers two principal disability programs that in Fiscal Year (FY) 2020 will together pay \$210 billion in benefits to as many as 18 million Americans.⁴⁷ An applicant to the more significant of these programs must demonstrate that his physical and/or mental impairments render him unable to work.⁴⁸ If successful, a typical claimant will receive about \$270,000 in lifetime benefits, plus Medicare coverage.⁴⁹

A claimant first files for disability benefits with a state-run office supervised by SSA.⁵⁰ A non-attorney “claims representative” decides whether the person meets the requirements for disability, based on the written application and supporting records. After requesting reconsideration, an unsuccessful claimant can further appeal to an ALJ for *de novo* review. The ALJ “develops the record,” ensuring that the claimant’s file includes complete medical records and other supporting evidence, before holding a hearing. Afterward, the ALJ prepares instructions and sends them to a “decision writer,” typically an attorney, to draft the decision. If the ALJ denies the claim, the claimant can appeal to SSA’s Appeals Council, an appellate body within the agency. Its administrative appeals judges, assisted by paralegal or attorney analysts, apply a “substantial evidence” standard of review to decide appeals.⁵¹ If the claimant loses again, he can seek judicial review in a federal district court.

SSA is the largest employer of federal ALJs, with roughly 1,600 ALJs working in about 170 hearing offices around the country.⁵² Until very

46. *See, e.g., Mathews v. Eldridge*, 424 U.S. 319, 342 (1976) (recognizing that an erroneous deprivation of Social Security benefits can cause a “significant” “hardship”); *Veterans for Common Sense v. Shinseki*, 678 F.3d 1013, 1035 (9th Cir. 2012) (en banc) (recognizing the “compelling interest” a veteran has in disability benefits and that an erroneous deprivation “can be devastating”).

47. *See* Soc. Sec. Admin., FY 2020 Congressional Justification 4 (2019), <https://perma.cc/R7DE-5MKC> (describing Disability Insurance and Supplemental Security Income payments).

48. 42 U.S.C. § 423(d)(2)(A) (2018).

49. JONAH GELBACH & DAVID MARCUS, A STUDY OF SOCIAL SECURITY LITIGATION IN THE FEDERAL COURTS 47-48 (2016).

50. For the process details discussed in this paragraph, see *id.* at 9, 16-30, 56-57.

51. SOC. SEC. ADMIN., HEARINGS, APPEALS, AND LITIGATION LAW MANUAL I-3-3-4 (2017), <https://perma.cc/662J-AG84>.

52. Soc. Sec. Admin., *supra* note 47, at 200 (ALJs); *Hearing Office Locator*, SOC. SECURITY ADMIN., <https://perma.cc/2GZX-GRLA> (archived Oct. 19, 2019) (hearing offices).

recently, ALJs secured jobs through a nonpolitical process administered by the Office of Personnel Management (OPM).⁵³ In 2018, however, the Supreme Court cast doubt on this system's constitutionality,⁵⁴ and in July 2018, President Trump asserted his power to appoint ALJs independent of the OPM process.⁵⁵ ALJs enjoy qualified decisional independence, as guaranteed by the Administrative Procedure Act (APA), and they are exempt from performance evaluation.⁵⁶ They can only be removed from office for cause,⁵⁷ through a protracted process that involves a right to be heard before another agency adjudicator and ultimately judicial review.⁵⁸

2. Immigration adjudication

IJs work for EOIR in the Department of Justice. They decide several types of cases, all of which involve efforts by immigrants to stay in the United States.⁵⁹ An immigrant appears before an IJ in one of two ways. The Department of Homeland Security (DHS) can issue the immigrant a notice to appear if it believes the immigrant must leave. The immigrant then must attempt to convince the IJ that she has a right to remain. Alternatively, an asylum seeker can request an IJ hearing if an asylum officer denies an asylum application. The IJ is supposed to help the immigrant develop the record before her hearing on the merits of her claim for relief. After the hearing, at which a DHS attorney represents the United States, the IJ prepares a decision, sometimes with staff attorney assistance. Either DHS or the immigrant can appeal, first to EOIR's Board of Immigration Appeals (its internal appellate body), and then to one of the regional courts of appeals.⁶⁰

The national ranks of IJs have expanded quickly, with 192 hired since January 2017.⁶¹ The country's sixty-three immigration courts now employ 424

53. See BEERMANN & MASCOTT, *supra* note 29, at 1-2.

54. See *Lucia v. SEC*, 138 S. Ct. 2044, 2051, 2055 (2018).

55. Exec. Order No. 13,843, 83 Fed. Reg. 32,755, 32,755 (July 13, 2018).

56. See 5 U.S.C. § 4301(2)(D) (2018) (exempting ALJs from the definition of "employee" subject to performance appraisal); *Nash v. Califano*, 613 F.2d 10, 15 (2d Cir. 1980) (recognizing that ALJs have "a qualified right of decisional independence").

57. 5 U.S.C. § 7521(a).

58. *Id.* §§ 7521(a), 7703(a)(1).

59. See U.S. Dep't of Justice, Q2 Immigration Court Statistics for Fiscal Year 2018 (FY18), at 1 n.1 (2018), <https://perma.cc/L7AC-H8P4> (describing the types of cases IJs adjudicate).

60. See generally U.S. Dep't of Justice, Executive Office for Immigration Review: An Agency Guide (2017), <https://perma.cc/W9MV-UUBZ> (describing the immigration adjudication process within EOIR).

61. Alexander Stockton, *U.S. Immigration Courts Are Broken. These 5 Charts Show Why.*, VICE NEWS (June 10, 2019, 10:01 AM), <https://perma.cc/GCV2-TTLZ>.

IJs.⁶² The Attorney General hires IJs, with only minimal eligibility criteria limiting his choices.⁶³ IJs are subject to performance appraisals every two years, and a rating of “less than satisfactory” can result in the IJ’s removal.⁶⁴ An IJ can challenge a disciplinary action through a grievance process established in a collective bargaining agreement between the IJs’ union and EOIR.⁶⁵ Salaries for IJs approximate those of ALJs and VLJs.⁶⁶

3. Veterans’ benefits adjudication

The VA benefits system provides approximately \$90 billion per year to over 6.5 million veterans and their dependents.⁶⁷ Although claims for VA benefits involve education, home loans, burial expenses, and other matters, most seek compensation for injuries and disease incurred during military service.⁶⁸ Whereas a typical Social Security claimant is classified as either disabled or not disabled, a veteran’s disability can range from 0% to 100%, in ten-percentage-point increments for different impairments, ranging from “hearing loss” to “posttraumatic stress disorder.”⁶⁹ Monthly disability payments to veterans with no dependents can vary from \$140 to \$8,750.⁷⁰

A veteran seeking benefits first files a claim with one of the Veterans Benefits Administration’s fifty-six regional offices, which together completed close to 1.4 million claims in FY 2017.⁷¹ Non-attorney staff help the veteran assemble necessary records and then make a decision based on the

62. *Id.*; *Office of the Chief Immigration Judge*, U.S. DEP’T JUST., <https://perma.cc/3F3Z-DRGB> (last updated Feb. 21, 2019).

63. For requirements, see *Immigration Judge*, U.S. DEP’T JUST., <https://perma.cc/M86M-6X3Y> (last updated June 9, 2017).

64. Labor Agreement Between the National Association of Immigration Judges and USDOJ, Executive Office for Immigration Review arts. 22.2, 22.11 (2018), <https://perma.cc/4XPV-WRHH>.

65. *Id.* arts. 8-10.

66. Compare EXEC. OFFICE FOR IMMIGRATION REVIEW, U.S. DEP’T OF JUSTICE, 2018 IMMIGRATION JUDGE PAY RATES (2018), <https://perma.cc/ASR3-BPYG> (IJs), with U.S. OFFICE OF PERS. MGMT., 2018 LOCALITY RATES OF PAY: ADMINISTRATIVE LAW JUDGES (2018), <https://perma.cc/53BQ-NHWH> (ALJs), and *Job Announcement: Veterans Law Judge*, USAJOBS, <https://perma.cc/483Z-9D3B> (archived Dec. 10, 2019) (VLJs).

67. See 3 U.S. DEP’T OF VETERANS AFFAIRS, FY 2019 BUDGET SUBMISSION, at VBA-51 to -52 (2018).

68. See *id.* at VBA-52; 2016 VETERANS BENEFITS ADMIN. ANN. BENEFITS REP. 7-11.

69. See 38 C.F.R. pt. 4, app. C (2019); VETERANS BENEFITS MANUAL §§ 3.61, 5.2.8 (LexisNexis 2019); *Benefit Rates*, U.S. DEP’T VETERANS AFF., <https://perma.cc/BQ3L-WB6W> (archived Oct. 19, 2019).

70. *2018 COLA Rates*, U.S. DEP’T VETERANS AFF., <https://perma.cc/BE9Z-ZU26> (last updated Aug. 21, 2017).

71. 3 U.S. DEP’T OF VETERANS AFFAIRS, *supra* note 67, at VBA-53, GenAd-319.

Due Process and Mass Adjudication
72 STAN. L. REV. 1 (2020)

documentary record. If dissatisfied, the veteran can file a “notice of disagreement,”⁷² which obliges the regional office to take a second look and issue a “statement of the case.”⁷³ If the veteran continues to disagree with the decision, he can appeal to BVA in Washington, D.C., for de novo review.⁷⁴ There, BVA assists the veteran to make sure that the VA has properly developed the record, ensuring that the veteran’s claim includes all relevant documentary evidence.⁷⁵ Unless the veteran requests a hearing, a staff attorney will draft a decision before the VLJ considers a case’s merits. The VLJ will review the record and the draft, and then sign the decision once it is ready.⁷⁶ A veteran can appeal an unfavorable decision to the Court of Appeals for Veterans Claims (CAVC),⁷⁷ an Article I court, and from there to the Federal Circuit.⁷⁸

BVA has grown rapidly, with its VLJ corps expanding from 56 in 2006 to 92 by 2018.⁷⁹ Founded in 1933,⁸⁰ BVA predates the APA, and thus VLJs are not formally ALJs. The VA Secretary appoints them with the President’s approval and based upon the BVA Chairman’s recommendations.⁸¹ VLJs go through

72. 38 U.S.C. § 7105 (2018). The Veterans Appeals Improvement and Modernization Act of 2017 made significant changes to the VA appeals process, effective February 19, 2019. See Pub. L. No. 115-55, 131 Stat. 1105 (codified as amended in scattered sections of 38 U.S.C.); see also VA Claims and Appeals Modernization, 84 Fed. Reg. 138 (Jan. 18, 2019) (codified at 38 C.F.R. pts. 3, 8, 14, 19, 20, 21 (2019)). The description of the VA appeal system provided here is of the system as it existed prior to February 19, 2019. However, no substantive changes were made to BVA’s standard of review and the workload remains significant. Under the new system, BVA projects the number of cases it receives will more than double in 2019 and 2020, as compared to 2018. See 2018 BOARD VETERANS’ APPEALS ANN. REP. 24, 27.

73. See *File a VA Disability Appeal*, U.S. DEP’T VETERANS AFF. (capitalization altered), <https://perma.cc/88U5-L6D8> (last updated Sept. 27, 2019).

74. See 38 U.S.C. § 7104.

75. See, e.g., *McGee v. Peake*, 511 F.3d 1352, 1357 (Fed. Cir. 2008).

76. 2001 BOARD VETERANS’ APPEALS REP. CHAIRMAN 14; VETERANS BENEFITS MANUAL § 13.23 (LexisNexis 2019).

77. 38 U.S.C. § 7252. Congress created CAVC and provided for judicial review of BVA decisions after Article III courts, prompted by *Goldberg v. Kelly* and other developments, increasingly disregarded statutory language purporting to limit their review authority. See H.R. REP. NO. 100-963, pt. 1, at 17-22 (1988). CAVC’s creation, in turn, prompted a judicialization of review at BVA. See James D. Ridgway, *The Veterans’ Judicial Review Act Twenty Years Later: Confronting the New Complexities of the Veterans Benefits System*, 66 N.Y.U. ANN. SURV. AM. L. 251, 273-75 (2010).

78. 38 U.S.C. § 7292.

79. 2018 BOARD VETERANS’ APPEALS, *supra* note 72, at 36; 2006 BOARD VETERANS’ APPEALS REP. CHAIRMAN 2.

80. 1992 BOARD VETERANS’ APPEALS REP. CHAIRMAN 3.

81. 38 U.S.C. § 7101A(a)(1).

performance reviews every one to three years to obtain recertification.⁸² In other respects, though, they resemble ALJs: They operate on the same pay scale as ALJs, they enjoy “for cause” removal protection, and they have a right to an agency hearing and ultimately judicial review to challenge disciplinary actions taken against them.⁸³

B. The Caseload Crisis and the Quantity-Quality Tradeoff

At the end of 1998, the average SSA ALJ had 326 hearings pending.⁸⁴ That figure more than doubled to 723 by the end of 2016.⁸⁵ On average, a disability benefits claimant in 2012 waited 353 days for an ALJ to decide her case. By the end of 2017, the average wait time had risen to 605 days.⁸⁶ Over 10,000 claimants died awaiting a hearing in 2017 alone.⁸⁷

Caseloads for IJs have likewise exploded. From 1998 to 2018, the number of deportation cases pending in immigration courts grew nearly sixfold, from 129,505 to 768,257, while the number of IJs increased by only about two-thirds.⁸⁸ At their present pace, one 2018 estimate suggested IJs would take 2.6 years to clear the then-existing backlog, even if the government initiated no cases in the interim.⁸⁹ But filings, in fact, have increased at staggering rates. In FY 2017, a record 295,222 new cases were filed in immigration courts; FY 2019 topped that figure by 148,507 cases.⁹⁰ Although times vary considerably by facility, immigrants in many courts now wait more than 1,300 days for a hearing,⁹¹ and in some courts that delay can be as long as six years.⁹²

82. *Id.* § 7101A(c), (d).

83. *See* 5 U.S.C. § 7513 (2018); 38 U.S.C. § 7101A(e)(2); 1995 BOARD VETERANS’ APPEALS REP. CHAIRMAN 7.

84. 2000 SOC. SECURITY ADMIN. ANN. STAT. SUPPLEMENT 117 tbl.2.F8.

85. 2018 SOC. SECURITY ADMIN. ANN. STAT. SUPPLEMENT 2.79 tbl.2.F8. An increase in ALJ hiring lowered this number to 661 per ALJ in 2017. *See id.*

86. Jackie Walorski, Opinion, *Reforms Are Needed to Fix Social Security Disability Insurance Backlog*, HILL (Dec. 18, 2017, 12:20 PM EST), <https://perma.cc/EA5Y-4E4W>.

87. *Id.*

88. *Backlog of Pending Cases in Immigration Courts as of August 2019*, TRAC IMMIGR., <https://perma.cc/D3FX-54W8> (archived Oct. 19, 2019); Nat’l Ass’n of Immigration Judges, *supra* note 6, at 3.

89. Miguel de Figueiredo et al., *Swift Injustice*, SLATE (Apr. 18, 2018, 3:51 PM), <https://perma.cc/H7DX-LCE4>. The situation has likely worsened, as the backlog of active cases recently crossed the one million mark. *Immigration Court’s Active Backlog Surpasses One Million*, TRAC IMMIGR. (Sept. 18, 2019), <https://perma.cc/BR54-KQS7>.

90. U.S. Dep’t of Justice, Executive Office for Immigration Review Adjudication Statistics (2019), <https://perma.cc/WQ5L-VP8V>.

91. *Crushing Immigration Judge Caseloads and Lengthening Hearing Wait Times*, TRAC IMMIGR. (Oct. 25, 2019), <https://perma.cc/E5B6-JHQP>.

92. Nat’l Ass’n of Immigration Judges, *supra* note 6, at 3.

Due Process and Mass Adjudication
72 STAN. L. REV. 1 (2020)

BVA's caseload has exploded over the past decade, recently doubling over three years between 2015 and 2018.⁹³ At the end of FY 2018, BVA had, on average, 1,821 cases pending per VLJ.⁹⁴ If its caseload continues to grow as projected, each VLJ would, on average, have a docket of 4,071 cases by FY 2024, assuming no increase in staffing beyond BVA's latest budget request.⁹⁵ In FY 2017, veterans' average wait times exceeded five years from the day they filed their appeals until a BVA decision.⁹⁶ Despite recent legislation to reform the appeals system, the VA projects BVA's backlog will more than double between 2018 and 2024.⁹⁷ Thousands of veterans die while their appeals languish.⁹⁸ Because BVA remands half of its cases for further development,⁹⁹ veterans commonly "find themselves trapped for years in a bureaucratic labyrinth, plagued by delays and inaction," as a judge on the Federal Circuit recently lamented.¹⁰⁰ Worsening backlogs threaten to exacerbate this "churn."¹⁰¹

While the backlogs have some distinct causes across the agencies,¹⁰² each agency has responded with similar measures. First, each agency has added officials at a rapid clip.¹⁰³ Beyond that, each has ratcheted up expectations for

93. Ridgway, *supra* note 2, at 3.

94. See U.S. DEP'T OF VETERANS AFFAIRS, COMPREHENSIVE PLAN FOR PROCESSING LEGACY APPEALS AND IMPLEMENTING THE MODERNIZED APPEALS SYSTEM: PUBLIC LAW 115-55, SECTION 3; AUGUST 2018 UPDATE 20 (2018), <https://perma.cc/34UJ-KLCB> (listing the number of VLJs as ninety-one); *id.* at 24 (describing BVA inventory).

95. *Id.* at 20, 24; 3 U.S. DEP'T OF VETERANS AFFAIRS, FY 2020 BUDGET SUBMISSION, at BVA-265 tbl. (2019), <https://perma.cc/Z4LZ-F4QP> (requesting ninety-eight VLJs for FY 2020).

96. 2017 BOARD VETERANS' APPEALS, *supra* note 1, at 25.

97. See U.S. DEP'T OF VETERANS AFFAIRS, *supra* note 94, at 24.

98. See OFFICE OF INSPECTOR GEN., U.S. DEP'T OF VETERANS AFFAIRS, 16-01750-79, VETERANS BENEFITS ADMINISTRATION: REVIEW OF TIMELINESS OF THE APPEALS PROCESS, at iv-v (2018), <https://perma.cc/74BV-D6T5> (reporting that, during the first quarter of FY 2016, approximately 1,600 veterans died while their appeals were pending).

99. 2018 BOARD VETERANS' APPEALS, *supra* note 72, at 32.

100. *Martin v. O'Rourke*, 891 F.3d 1338, 1349 (Fed. Cir. 2018) (Moore, J., concurring).

101. For the use of the term "churn" to describe the cycle of appeals, see, for example, Nicholas B. Holtz, *The Churn of Cases Within VA's Appeals Process*, VETERANS L.J., Spring 2015, at 1.

102. On possible causes, see, for example, Mark Duggan, *The Labor Market Effects of the VA's Disability Compensation Program*, STAN. INST. FOR ECON. POL'Y RES. POL'Y BRIEF, Nov. 2014, at 1, 2-3 (VA); Lorelei Laird, *Legal Logjam*, A.B.A. J., Apr. 2017, at 52, 55-56 (EOIR); and T.J. Sutcliffe, Guest Editorial, *Social Security's Disability Hearings Backlog: A National Crisis*, 41 HEALTH & SOC. WORK 215, 216 (2016) (SSA).

103. 2018 SOC. SECURITY ADMIN., *supra* note 85, at 2.79 tbl.2.F8 (noting the addition of more than one hundred new ALJs); 2017 BOARD VETERANS' APPEALS, *supra* note 1, at 14 (describing the addition of twenty-six new VLJs); Stockton, *supra* note 61 (describing the rapid increase in IJ hiring).

how quickly judges decide cases. When it first set a case completion goal in 1975, SSA asked ALJs to decide twenty-six cases per month.¹⁰⁴ In March 2017, SSA set this figure at fifty decisions per month.¹⁰⁵ In April 2018, the Attorney General threatened to issue subpar performance ratings to any IJ deciding fewer than 700 cases per year¹⁰⁶—about twenty more than the average IJ decided annually from 2011 to 2016.¹⁰⁷ BVA expects VLJs to decide twenty to thirty cases per week,¹⁰⁸ a threshold that in FY 2018 helped the agency render 85,288 decisions, or 32,627 more than the year before.¹⁰⁹ Before, “the line judges were doing 750 [to] 800 [cases],” one official told us, but “it’s now up to 1,000 cases a year.”¹¹⁰ Another official described the case completion goal as “far higher now than it was five years ago.”¹¹¹

These caseload pressures pose a systemic threat. Conventional wisdom holds that an uptick in the pace of decisionmaking degrades decisional quality.¹¹² This quantity-quality tradeoff has a straightforward logic: The faster a decisionmaker has to work, the more she is likely to err.¹¹³

Many of those involved with high-volume agency adjudication, going back decades, accept this conventional wisdom.¹¹⁴ More than 70% of several

104. DONNA PRICE COFER, JUDGES, BUREAUCRATS, AND THE QUESTION OF INDEPENDENCE: A STUDY OF THE SOCIAL SECURITY ADMINISTRATION HEARING PROCESS 81 (1985).

105. SOC. SEC. ADMIN., 2017 UPDATED COMPASSIONATE AND RESPONSIVE SERVICE (CARES) AND ANOMALY PLAN 14 (2017), <https://perma.cc/DNL8-UUAN>.

106. Nick Miroff, *Trump Administration, Seeking to Speed Deportations, to Impose Quotas on Immigration Judges*, WASH. POST (Apr. 2, 2018, 4:16 PM PDT), <https://perma.cc/JA42-YSLB>.

107. See Liz Robbins, *In Immigration Courts, It Is Judges vs. Justice Department*, N.Y. TIMES (Sept. 7, 2018), <https://perma.cc/Y7CM-DFYH>.

108. Telephone Interview with Interview Subject No. 4, *supra* note 5.

109. 2018 BOARD VETERANS’ APPEALS, *supra* note 72, at 8.

110. Telephone Interview with Interview Subject No. 2 (Aug. 29, 2018).

111. Telephone Interview with Interview Subject No. 4, *supra* note 5.

112. See, e.g., *Heckler v. Day*, 467 U.S. 104, 113-15 (1984); LEWIN GRP., INC. ET AL., EVALUATION OF SSA’S DISABILITY QUALITY ASSURANCE (QA) PROCESSES AND DEVELOPMENT OF QA OPTIONS THAT WILL SUPPORT THE LONG-TERM MANAGEMENT OF THE DISABILITY PROGRAM 17 (2001) (noting “a widespread belief in a trade-off between accuracy and productivity” among SSA adjudicators); Stuart L. Lustig et al., *Inside the Judges’ Chambers: Narrative Responses from the National Association of Immigration Judges Stress and Burnout Survey*, 23 GEO. IMMIGR. L.J. 57, 65-66 (2008) (reporting perceptions about this tradeoff among IJs).

113. See *Ass’n of Admin. Law Judges v. Colvin*, 777 F.3d 402, 404 (7th Cir. 2015); cf. *C.J.L.G. v. Barr*, 923 F.3d 622, 636-37 (9th Cir. 2019) (en banc) (Paez, J., concurring) (describing IJs’ quotas and insisting that “this enormous workload” must downgrade the quality of immigration adjudication).

114. Agency officials have on occasion denied that an increased pace of decisionmaking comes at the expense of decisional accuracy. See, e.g., Gerald K. Ray & Jeffrey S. Lubbers, *A Government Success Story: How Data Analysis by the Social Security Appeals Council* (with a footnote continued on next page

hundred ALJs surveyed in 1978, for example, reported that production pressures caused decisional quality to suffer.¹¹⁵ SSA studied the quantity-quality relationship in 2012 and concluded that “[a]s ALJ production increases, the general trend for decisional quality is to go down.”¹¹⁶ Among 395 BVA attorneys informally surveyed in 2018, most agreed that their production standard “results in a decrease in the quality of decisions for veterans, and substantially increases the likelihood of errors.”¹¹⁷

* * *

Much rides on decisional quality. An erroneous denial of benefits can subject a veteran or disabled worker to extreme hardship while he slogs through a frustratingly slow appeals process. “Delay, deny, hope they die” has become a slogan of sorts for cynical veterans stuck in the VA process.¹¹⁸ Errors coupled with long delay times can be a “death sentence” for Social Security disability benefits claimants, as a district judge recently noted.¹¹⁹ It takes little to imagine the worst for an asylum applicant, facing persecution at home, who is wrongly denied relief.¹²⁰ Poor-quality adjudication also has systemic consequences, for the federal budget, the federal courts, and the just and accurate implementation of immigration policy.

Push from the Administrative Conference of the United States) Is Transforming Social Security Disability Adjudication, 83 GEO. WASH. L. REV. 1575, 1606 (2015) (describing “quality improvement” at SSA’s Appeals Council notwithstanding a “significant increase in dispositions”).

115. See STAFF OF H. SUBCOMM. ON SOC. SEC. OF THE COMM. ON WAYS & MEANS, 96TH CONG., SOCIAL SECURITY ADMINISTRATIVE LAW JUDGES: SURVEY AND ISSUE PAPER 6, 32 (Comm. Print 1979).
116. STAFF OF H. COMM. ON OVERSIGHT & GOV’T REFORM, 113TH CONG., MISPLACED PRIORITIES: HOW THE SOCIAL SECURITY ADMINISTRATION SACRIFICED QUALITY FOR QUANTITY IN THE DISABILITY DETERMINATION PROCESS 43 (2014) (alteration in original) (quoting an SSA official).
117. Letter from Douglas E. Massey, President, Am. Fed’n of Gov’t Emps. Local 17, to Cheryl Mason, Chairman, Bd. of Veterans’ Appeals 3-4 (Apr. 26, 2018) (on file with authors).
118. See, e.g., WALTER C. PINKNEY SR., DELAY, DENY, HOPE THEY DIE (2015); Susan Thompson, *Navigating the Department of Veterans Affairs with Invisible Wounds: How to Overcome the Stigma of “Delay, Deny, & Hope You Die,”* CIR POL’Y BRIEF, Apr. 2012, at 1; Craig M. Wax, VA to Vets: “Delay, Deny, Wait till They Die,” WASH. TIMES (Dec. 1, 2015), <https://perma.cc/HH23-PQJL>.
119. See *Boatner v. Berryhill*, No. 3:16-CV-243, 2018 WL 2191804, at *10 (S.D. Miss. May 11, 2018) (quoting Terrence McCoy, *597 Days. And Still Waiting*, WASH. POST (Nov. 20, 2017), <https://perma.cc/2L6H-8F44>).
120. See, e.g., Kevin Sieff, *When Death Awaits Deported Asylum Seekers*, WASH. POST (Dec. 26, 2018), <https://perma.cc/RZ53-QH7L>.

II. Systemic Quality, Constitutional Obligation, and Internal Administrative Law

If caseload pressures have indeed placed high-volume agency adjudication on a precipice, a plunge could have enormous policy and political implications. A plunge will also have legal ramifications, because systemic accuracy in high-volume agency adjudication is a constitutional obligation. A failure to achieve an adequate threshold of decisional quality can trigger a due process challenge. One response to the looming crisis of decisional quality could thus take the form of an order, imposed by a federal judge, instructing the agency to change how it does business. Under present doctrine, however, this sort of externally imposed redress is unlikely.

If decisional quality risks becoming an “underenforced constitutional norm[]” in the federal courts,¹²¹ internal administrative law may fill the void. After examining the limits of external sources of agency governance, this Part introduces quality assurance as a key component in an internally administered effort to improve adjudicator performance.

A. The Limits of External Law

We examine here the sources of external law that govern the accuracy of mass adjudication. We show that the actual operation of the doctrine sharply limits the ability of courts to correct systematic sources of error.

1. Accuracy, constitutional obligation, and remedial mismatch

An agency’s obligation to meet an adequate threshold of decisional quality comes out of the Due Process Revolution. In *Goldberg v. Kelly*, the revolution’s trigger,¹²² the Supreme Court declared that due process requires a welfare agency to afford beneficiaries certain procedural rights, including a hearing, before it can terminate benefits.¹²³ This “pre-termination hearing has one function only,” the Court explained: “to produce an initial determination of the validity of the . . . grounds for discontinuance . . . in order to protect a recipient against an *erroneous termination*.”¹²⁴ The Court in *Mathews v. Eldridge* recast *Goldberg* as its now-famous balancing test for the adjudication of due process

121. Lawrence Gene Sager, *Fair Measure: The Legal Status of Underenforced Constitutional Norms*, 91 HARV. L. REV. 1212, 1226 (1978).

122. 397 U.S. 254 (1970); see also Jerry L. Mashaw, *The Supreme Court’s Due Process Calculus for Administrative Adjudication in Mathews v. Eldridge: Three Factors in Search of a Theory of Value*, 44 U. CHI. L. REV. 28, 29 (1976) (noting that a “due process revolution” followed *Goldberg*).

123. *Goldberg*, 397 U.S. at 261.

124. *Id.* at 267 (emphasis added).

challenges to agency decisionmaking.¹²⁵ Whether the agency offers constitutionally adequate procedures when it threatens liberty or property interests depends in important part on “the risk of an erroneous deprivation of such interest through the procedures used.”¹²⁶ If the private interest in the enjoyment of the benefit at stake is strong enough, and if the procedures the agency offers to decide entitlements to that benefit err too frequently, then a court can order the agency to offer enhanced procedural rights.¹²⁷

Together, *Goldberg* and *Mathews* have three germane legacies. First, they cast the quality of agency adjudication in terms of accuracy. To be fair, this understanding of quality may be unduly narrow.¹²⁸ Justice Brennan, who wrote for the Court in *Goldberg*, insisted that the opinion also addresses the question of “[w]hether the government treats its citizens with dignity,”¹²⁹ consistent with the view that stresses the importance of fair process for “generating the feeling, so important to a popular government, that justice has been done.”¹³⁰ As Jerry Mashaw argued, a metric for quality decisionmaking that attends solely to accuracy “conceives of the values of procedure too narrowly,”¹³¹ and *Mathews*’s emphasis on better decisional “technique” is “incomplete” as being “unresponsive to the full range of concerns embodied in the due process clause.”¹³² But, as Mashaw and others have conceded, *Mathews* cast due process concerns narrowly in terms of accuracy.¹³³ As such, rightly or

125. See *Mathews v. Eldridge*, 424 U.S. 319, 334-35 (1976).

126. *Id.* at 335.

127. See *id.* Whether enhanced procedural rights are in order also depends on whether the government’s countervailing interest is compelling. See *id.*; cf. *Hamdi v. Rumsfeld*, 542 U.S. 507, 532-33 (2004) (plurality opinion) (applying the *Mathews* balancing test to require additional procedural protections).

128. See, e.g., LEWIN GRP. ET AL., *supra* note 112, at 3 (noting that SSA has emphasized accuracy in its definition of quality but suggesting that quality defined in other literature includes a “broad range of . . . attributes”).

129. William J. Brennan, Jr., *Reason, Passion, and “The Progress of the Law,”* 10 CARDOZO L. REV. 3, 21-22 (1988).

130. *Joint Anti-Fascist Refugee Comm. v. McGrath*, 341 U.S. 123, 171-72 (1951) (Frankfurter, J., concurring).

131. Mashaw, *supra* note 122, at 48.

132. *Id.* at 30.

133. See, e.g., *id.* at 48; see also *Addington v. Texas*, 441 U.S. 418, 425 (1979) (citing *Mathews* for the claim that “the function of legal process is to minimize the risk of erroneous decisions”). But see Gary Lawson et al., “Oh Lord, Please Don’t Let Me Be Misunderstood!”: *Rediscovering the Mathews v. Eldridge and Penn Central Frameworks*, 81 NOTRE DAME L. REV. 1, 22-23 (2005) (conceding the prevailing view of *Mathews* but arguing that *Mathews* can be read differently).

wrongly, accuracy continues to be the linchpin of due process in agency adjudication.¹³⁴

Second, *Goldberg* and *Mathews* emphasize that a system's overall accuracy, not the accuracy of an individual decision, matters to the due process calculus.¹³⁵ In most instances, the adjudicative process does not offend due process if it errs in one case or another.¹³⁶ The procedures courts can order as remedies for due process violations therefore have as their "sole goal" systemic accuracy, or accuracy for the category of decisionmaking at issue.¹³⁷ *Goldberg's* and *Mathews's* third legacy is the set of additional procedural protections that courts can order if a system errs too often. The logic is that these enhanced procedural rights will reduce errors in adjudication.

Critics of the Due Process Revolution have long questioned whether individual procedural rights can ensure that an agency meets its constitutional obligation to systemic accuracy. To the extent that such rights can improve the overall accuracy of an agency's decisionmaking, as a practical matter they do so only if unrelated individuals collectively exercise their rights in significant enough numbers and with some degree of unintended coordination. This happy result is unlikely.¹³⁸ Also, the population that ostensibly benefits from

134. Cf. LEWIN GRP. ET AL., *supra* note 112, at iv-vii (linking accuracy to quality); Jerry L. Mashaw, *Conflict and Compromise Among Models of Administrative Justice*, 1981 DUKE L.J. 181, 185 ("From the perspective of bureaucratic rationality, administrative justice is accurate decision-making carried on through processes that take account of costs."); *id.* at 207 (associating *Mathews* with a bureaucratic rationality model of administrative justice).

135. See Jason Parkin, *Due Process Disaggregation*, 90 NOTRE DAME L. REV. 283, 292-98 (2014).

136. See *Mathews v. Eldridge*, 424 U.S. 319, 344 (1976); see also *Parham v. J.R.*, 442 U.S. 584, 615 (1979) (citing *Mathews* for the insistence that courts judge the adequacy of process based on the generality of cases and not on error in a specific one).

137. See Mashaw, *supra* note 122, at 48; see also Richard H. Fallon, Jr., *Some Confusions About Due Process, Judicial Review, and Constitutional Remedies*, 93 COLUM. L. REV. 309, 336-37 (1993) (describing *Mathews* as requiring procedures that achieve a "tolerable average level of accuracy"); Gillian E. Metzger, *The Constitutional Duty to Supervise*, 124 YALE L.J. 1836, 1865 (2015) (describing *Mathews's* "decidedly systemic and managerial cast"); Robert L. Rabin, *Federal Regulation in Historical Perspective*, 38 STAN. L. REV. 1189, 1314 (1986) (describing *Mathews* as concerned with "potential error costs" to the group of potential recipients).

138. See, e.g., JOEL F. HANDLER, *THE CONDITIONS OF DISCRETION: AUTONOMY, COMMUNITY, BUREAUCRACY* 32-40 (1986) (describing nonlegal, nonprocedural barriers that prevent individuals from protecting their property interests); David Hausman, *The Failure of Immigration Appeals*, 164 U. PA. L. REV. 1177, 1214 (2016) (explaining how a denial of continuances can prevent immigrants from appealing cases); Mashaw, *supra* note 17, at 775-76 ("[T]he purposes, necessary modes of operation, and clientele of social welfare programs . . . severely limit the value of procedural safeguards and appellate checks . . ."); Jason Parkin, *Adaptable Due Process*, 160 U. PA. L. REV. 1309, 1331 (2012) (summarizing critiques of the "fair hearing" requirement for its efficacy at protecting property interests for vulnerable populations); William H. Simon, *Legality*,
footnote continued on next page

enhanced procedural rights may lack the resources or wherewithal to exercise these rights in large enough numbers.¹³⁹ Appeals, to name one procedure recently lauded as a sort of cure-all for erroneous decisionmaking,¹⁴⁰ can be “highly and mysteriously selective,”¹⁴¹ and may depend on factors wholly unrelated to accuracy.¹⁴² IJs’ decisions to deny continuances to immigrants seeking counsel and the government’s decision to detain immigrants, for instance, both depress appeals.¹⁴³ Appellate dockets thus may neither reflect actual error patterns nor lead an agency to understand systematic sources of error.

An individual’s use of an accuracy-promoting procedure, without more, does nothing to promote systemic accuracy.¹⁴⁴ A single appeal may correct a single error, for example, but it does not oblige the agency or a reviewing court “to determine whether” the appeal represents “an isolated problem or the tip of an iceberg of similar but unappealed cases.”¹⁴⁵ A critical mass of appeals over time could, in principle, reveal entrenched pathologies in agency decisionmaking, even if the selection of cases for appellate review is imperfect.¹⁴⁶ But in practice, the agency must make some coordinated, managerial effort to gather and analyze data on appeals patterns and then act on what this information reveals. On its own, the mere provision of a right to appeal does not require or enable the agency to identify and correct systemic errors.

Bureaucracy, and Class in the Welfare System, 92 YALE L.J. 1198, 1248-49 (1983) (describing the inefficacy of appeals due to claimants’ limitations).

139. See HANDLER, *supra* note 138, at 22-23; Mashaw, *supra* note 17, at 775-76; William H. Simon, *The Organizational Premises of Administrative Law*, 78 LAW & CONTEMP. PROBS., nos. 1 & 2, 2015, at 61, 82-83.

140. See C.J.L.G. v. Sessions, 880 F.3d 1122, 1143-44 (9th Cir. 2018), *vacated en banc sub nom.* C.J.L.G. v. Barr, 923 F.3d 622 (9th Cir. 2019).

141. Mashaw, *supra* note 17, at 784.

142. See Hausman, *supra* note 138, at 1181; Steven Shavell, *The Appeals Process as a Means of Error Correction*, 24 J. LEGAL STUD. 379, 413 (1995). Even when litigants have imperfect information about errors, their selection of cases for appeal may select for errors. But the social costs of appeals under such conditions are likely to be high, in part because poorly chosen appeals will make it more difficult for appellate courts to recognize error. Shavell, *supra*, at 413.

143. BANKS MILLER ET AL., IMMIGRATION JUDGES AND U.S. ASYLUM POLICY 131-32 (2015); Hausman, *supra* note 138, at 1214.

144. See Mashaw, *supra* note 17, at 785; see also Simon, *supra* note 139, at 82 (noting that an agency “has no duty to follow up on systemically relevant information disclosed in adjudications”).

145. Mashaw, *supra* note 17, at 785.

146. See Shavell, *supra* note 142, at 413.

Finally, the provision of individual procedural rights may have systemic costs, including to systemic accuracy.¹⁴⁷ More errors and more appeals should produce more remands, and thus further increase caseload pressures impacting quality. A reviewing court, exposed to only a sliver of the agency's caseload, may impose a burdensome requirement on the agency that increases costs and delay dramatically. Requiring the VA to provide veterans seeking benefits with expert medical opinions, for instance, means a government-funded examination "probably better than every three seconds," as one official, skeptical of the Federal Circuit's supervision of VA decisionmaking, told us: "Right now, each time you breathe out, the VA just produced an expert medical opinion on a claim."¹⁴⁸ The result is that one claimant's, immigrant's, or veteran's decision to avail herself of procedural protections to avoid or correct an error may lead to more errors in other cases, with questionable benefits and potentially significant costs for supposed due process beneficiaries.¹⁴⁹

To be sure, strengthened procedural rights for individuals to exercise may improve systemic accuracy. A due process right to court-appointed and government-funded counsel, for example, would probably strengthen the overall quality of immigration adjudication.¹⁵⁰ This right for children in removal proceedings almost surely would do so.¹⁵¹ But, even if as-of-yet undeclared rights could improve systemic accuracy for different systems of adjudication, the likelihood that courts will find them lurking in the Due Process Clause is exceedingly low.¹⁵² Claims seeking additional procedural protections for social welfare beneficiaries have uniformly failed in the

147. William H. Simon, *The Rule of Law and the Two Realms of Welfare Administration*, 56 BROOK. L. REV. 777, 786 (1990); see JERRY L. MASHAW, *DUE PROCESS IN THE ADMINISTRATIVE STATE* 34 (1985).

148. Telephone Interview with Interview Subject No. 2, *supra* note 110.

149. See James D. Ridgway & David S. Ames, *Misunderstanding Chenery and the Problem of Reasons-or-Bases Review*, 68 SYRACUSE L. REV. 303, 316-24 (2018) (discussing how "reasons-or-bases" CAVC remands rarely result in the award of benefits).

150. Cf. Ingrid V. Eagly & Steven Shafer, *A National Study of Access to Counsel in Immigration Court*, 164 U. PA. L. REV. 1, 49-50 (2015) (describing the positive impact of access to counsel for immigrants).

151. See *C.J.L.G. v. Barr*, 923 F.3d 622, 634 (9th Cir. 2019) (en banc) (Paez, J., concurring).

152. See ADRIAN VERMEULE, *LAW'S ABNEGATION: FROM LAW'S EMPIRE TO THE ADMINISTRATIVE STATE* 123 (2016). In *Walters v. National Ass'n of Radiation Survivors*, for example, the Supreme Court found that a ten-dollar fee limitation on representation in front of the VA did not violate due process, in spite of the fact that constraints on legal representation may cause errors. See 473 U.S. 305, 319-34 (1985).

Supreme Court this century,¹⁵³ and to date immigrants have failed to obtain additional rights under even the most sympathetic of circumstances.¹⁵⁴

2. Jurisdiction channeling

Procedural due process is not the only external source of legal obligation concerned with decisional quality. The APA is another,¹⁵⁵ as are agency-specific statutes.¹⁵⁶ In theory, nothing in the substance of these laws would prevent them from evolving to support claims plaintiffs can bring for remedies more systemic in reach than can be achieved through individual procedural rights.¹⁵⁷ Such an evolution would fit neatly with the “duty to supervise” that Gillian Metzger argues the Constitution imposes on federal agencies.¹⁵⁸ But statutes governing disability benefits, immigration, and veterans’ benefits litigation mostly limit the federal courts to an appellate role concerned with the individual remediation of errors.¹⁵⁹ These jurisdiction-channeling provisions would steer claims alleging systemic problems with decisional quality to fora ill situated to receive them favorably.¹⁶⁰

In theory, a disability claimant, immigrant, or veteran could challenge some systemic inadequacy in agency decisionmaking as part of an individual appeal or petition for review. It stands to reason that an appellant or petitioner will rarely pursue, much less win, these remedies.¹⁶¹ Presumably a disability benefits claimant who appeals to a federal district court does so because she believes SSA denied her benefits erroneously. The agency may have inadequately supervised ALJs in a particular hearing office, and perhaps this

153. Merrill, *supra* note 14, at 44-45.

154. In *C.J.L.G. v. Sessions*, a three-judge panel of the Ninth Circuit held that children in removal proceedings lack a right to counsel. 880 F.3d 1122, 1145 (9th Cir. 2018). An en banc panel vacated the decision, but it did not decide the constitutional issue, remanding on statutory grounds instead. 923 F.3d at 627-29.

155. See 5 U.S.C. § 706(2), (2)(A) (2018) (authorizing a court to “set aside agency action” that is “arbitrary [and] capricious”).

156. See, e.g., 42 U.S.C. § 405(g) (2018) (authorizing a court to set aside a disability benefits determination not supported by “substantial evidence”).

157. *But see* MASHAW, *supra* note 19, at 187 (questioning whether anyone would have standing to bring a systemic challenge).

158. Metzger, *supra* note 137, at 1842.

159. See 8 U.S.C. § 1252(a)(5), (b)(9) (2018) (immigration); 38 U.S.C. §§ 511, 7252, 7292 (2018) (veterans’ benefits); 42 U.S.C. § 405(g)-(h) (disability benefits).

160. On jurisdiction channeling, see *Shalala v. Illinois Council on Long Term Care, Inc.*, 529 U.S. 1, 13 (2000). For a discussion of how a jurisdiction-channeling provision effectively prohibits a generic APA challenge to administrative adjudication, see *Martinez v. Napolitano*, 704 F.3d 620 (9th Cir. 2012).

161. See Gelbach & Marcus, *supra* note 43, at 1140-44 (documenting the evidentiary difficulty of showing systemic inaccuracy).

inadequate supervision made the error she alleges more likely. A systemic remedy, such as an order requiring better supervision, might indirectly benefit her. But she is more likely to benefit directly if she simply asks for the district court to correct the error in her individual case.

Legal and institutional hurdles stand in the way of an altruistic appellant or petitioner who might eschew a direct individualized remedy for an indirect, systemic one. Although its contours are contested, scope-of-remedy doctrine tends to limit a litigant to an injunction (or an injunction-like remedy) no broader than what is necessary to resolve the harm the litigant herself has experienced.¹⁶² A petitioner or appellant may proceed on behalf of a class of similarly situated people to broaden the remedy's scope.¹⁶³ But jurisdiction-channeling provisions steer litigants to appellate fora that do not typically handle class action-type challenges of this sort. Rule 23 of the Federal Rules of Civil Procedure does not apply in the courts of appeals,¹⁶⁴ and whatever other power they might have to certify classes of immigrant petitioners is uncertain and largely untested.¹⁶⁵ The Federal Circuit recently recognized CAVC's power to certify classes of veterans.¹⁶⁶ But neither CAVC nor the federal circuit courts can supervise a discovery process of the sort necessary to establish that agency practices produce systemic inaccuracies. Such evidence would have to be compiled in the agency before the appeal.¹⁶⁷ District courts, the appellate tribunals for SSA appeals, do have these procedural powers. But legal restrictions may limit their review to evidence in the record,¹⁶⁸ and exhaustion requirements may limit class membership to those who have pursued all internal agency appeals.¹⁶⁹

162. See *Trump v. Hawaii*, 138 S. Ct. 2392, 2424-29 (2018) (Thomas, J., concurring); DOUGLAS LAYCOCK, *MODERN AMERICAN REMEDIES: CASES AND MATERIALS* 246-47 (3d ed. 2002).

163. Samuel L. Bray, *Multiple Chancellors: Reforming the National Injunction*, 131 HARV. L. REV. 417, 475 (2017) (discussing class actions as the appropriate vehicle when injunctive relief for individual plaintiffs may be too narrow).

164. See FED. R. CIV. P. 1 ("These rules govern the procedure in all civil actions and proceedings in the United States district courts . . .").

165. Cf. Adam S. Zimmerman, *The Appellate Class Action* 5-6 (June 19, 2019) (unpublished manuscript) (on file with authors) (suggesting that under the All Writs Act, appellate courts may have claim-aggregating authority that resembles Rule 23).

166. *Ebanks v. Shulkin*, 877 F.3d 1037, 1040 (Fed. Cir. 2017); *Monk v. Shulkin*, 855 F.3d 1312, 1318 (Fed. Cir. 2017).

167. Cf. *Kyhn v. Shinseki*, 716 F.3d 572, 573-76 (Fed. Cir. 2013) (finding error where CAVC sought new evidence *sua sponte* to supplement the record).

168. See, e.g., *Pronti v. Barnhart*, 339 F. Supp. 2d 480, 497 (W.D.N.Y. 2004) (conceding that the court has no factfinding power in a Social Security appeal); *Small v. Barnhart*, 329 F. Supp. 2d 1272, 1277 (N.D. Ala. 2004) (noting that even a district court should adhere to a process of remanding for factfinding).

169. See *Bowen v. City of New York*, 476 U.S. 467, 482-83 (1986).

To be sure, jurisdiction-channeling provisions do permit some systemic challenges to aspects of administrative adjudication pursued outside the designated appeals or petitions process. For a plaintiff to file such a case directly in a district court, the systemic challenge must involve matters that are “collateral” to the merits of individual claims.¹⁷⁰ A challenge to an SSA policy treating in-kind loans as a form of income, for instance, was collateral because it was “not essentially a claim for benefits.”¹⁷¹ But the various tests courts use to determine when claims are indeed “collateral” tend to exclude those alleging systemic inaccuracy in adjudicator decisionmaking, on grounds that these allegations necessarily require examination of cases’ merits.¹⁷² Although the Supreme Court has at times interpreted jurisdiction-channeling provisions narrowly to ensure “meaningful” access to judicial review for claims of systemic maladministration,¹⁷³ the Court has more recently honored jurisdiction channeling so long as the provision does not “foreclose all judicial review” entirely.¹⁷⁴

The doctrine governing judicial review of administrative adjudication includes more nuance than we can discuss here, and we return to the possibility of systemic remedies imposed through structural reform litigation in Part V. For now, the point is simply that jurisdiction channeling reinforces the remedial limitations of procedural due process. It steers claims about decisional quality into venues that privilege individual error avoidance or correction over systemic reform. The result is a body of external law for administrative governance that presently offers little promise as a bulwark against a systemic decline in decisional quality.¹⁷⁵

170. See, e.g., *id.* at 483; *Weinberger v. Salfi*, 422 U.S. 749, 764 (1975); *J.E.F.M. v. Lynch*, 837 F.3d 1026, 1032 (9th Cir. 2016); *Aguilar v. U.S. Immigration & Customs Enf’t*, 510 F.3d 1, 11-15 (1st Cir. 2007); *Kildare v. Saenz*, 325 F.3d 1078, 1083 (9th Cir. 2003).

171. *Johnson v. Shalala*, 2 F.3d 918, 921-22 (9th Cir. 1993).

172. See, e.g., *J.E.F.M.*, 837 F.3d at 1032 (extending a jurisdiction-channeling provision to all claims that “arise from removal proceedings”); *Veterans for Common Sense v. Shinseki*, 678 F.3d 1013, 1025, 1027-28 (9th Cir. 2012) (en banc) (extending the jurisdiction-channeling provision for veterans’ benefits claims to challenges to decisions “that may affect” a “request for benefits”); *Kildare*, 325 F.3d at 1083 (limiting jurisdiction when cases about disability adjudication are “inextricably intertwined with . . . claims for benefits”).

173. See *Thunder Basin Coal Co. v. Reich*, 510 U.S. 200, 212-13 (1994); *McNary v. Haitian Refugee Ctr., Inc.*, 498 U.S. 479, 497-99 (1991).

174. *Elgin v. Dep’t of the Treasury*, 567 U.S. 1, 10 (2012).

175. Cf. *Metzger*, *supra* note 137, at 1870-71 (“Systemic administrative functioning [has been] denied constitutional relevance, and no constitutional claim can be made simply because a government agency or institution is inadequately managed or supervised.”).

B. Internal Administrative Law and Quality Assurance

In 1974, Mashaw argued for a different due process doctrine, emphasizing its “management side.”¹⁷⁶ Rather than only giving individuals procedures they can use to avoid or correct particular errors, due process should also require agencies to use “systematic management techniques [to] discover errors, identify their causes and implement corrective action.”¹⁷⁷ Mashaw quickly rethought his recommendation that courts enforce this version of due process with structural injunctions,¹⁷⁸ insisting in *Bureaucratic Justice* that such interventions would lie beyond judicial competence.¹⁷⁹ But Mashaw did not give up on a “right” to good administration. Rather, he turned away from due process and reforms imposed from outside the agency and toward an “internal law of administration” as a better means for achieving improvement in agency performance.¹⁸⁰

In recent years, Mashaw’s view that externally devised and imposed law, such as procedural due process and the APA, only controls decisionmaking at the margins has gained prominent purchase.¹⁸¹ “Internal administrative law”—the “internal rules and procedures, bureaucratic systems, and internal techniques of instruction, oversight and control of agency personnel”¹⁸²—matters far more to the day-to-day work of agency officials, judges, and employees.¹⁸³

Administrative law theorists have reinvigorated debates about internal administrative law as a promising vein to tap to “encourage consistency, predictability, and reasoned argument in agency decisionmaking.”¹⁸⁴ Deficits

176. See generally Mashaw, *supra* note 17 (describing due process’s “management side” and its remedial implications in terms of quality assurance).

177. *Id.* at 815-16.

178. See Jerry L. Mashaw, *How Much of What Quality? A Comment on Conscientious Procedural Design*, 65 CORNELL L. REV. 823, 834 (1980).

179. See MASHAW, *supra* note 19, at 226.

180. *Id.* at 213, 226; Merrill, *supra* note 14, at 53.

181. See, e.g., VERMEULE, *supra* note 152, at 217.

182. Jerry L. Mashaw, *Conclusion: The Inside Out Perspective: A First-Person Account*, in ADMINISTRATIVE LAW FROM THE INSIDE OUT, *supra* note 14, at 501, 502; see also Gillian E. Metzger & Kevin M. Stack, *Internal Administrative Law*, 115 MICH. L. REV. 1239, 1243-44, 1253 (2017) (describing the corpus of internal administrative law).

183. Cf. MASHAW, *supra* note 19, at 18-19 (critiquing the irrelevance of formal administrative law to the work of agency officials in SSA).

184. Metzger & Stack, *supra* note 182, at 1244; Walker, *supra* note 20, at 1624 (“Internal administrative law . . . has become a hot subfield in administrative law.”); see also Kristin A. Collins, *Bureaucracy as the Border: Administrative Law and the Citizen Family*, 66 DUKE L.J. 1727, 1731-32 (2017) (commenting on internal administrative law scholarship); Elizabeth Magill, Foreword, *Agency Self-Regulation*, 77 GEO. WASH. L. REV. 859, 860-61 (2009) (providing an overview of aspects of agency “self-regulation”); Jennifer Nou, *footnote continued on next page*

of institutional competence limit judicial capacity to devise doctrine that will succeed in ensuring that agencies balance all of the competing goals they have to serve.¹⁸⁵ Judicial review and external oversight are too infrequent and too abstract to ensure that granular realities of day-to-day decisionmaking align with legal requirements.¹⁸⁶ “Law’s abnegation” of the governance of agency adjudication reflects this reality, Adrian Vermeule argues: “The federal judicial system is not set up, not equipped, to engage in a sustained course of synoptic institutional engineering.”¹⁸⁷ Agencies, in contrast, have the practical expertise to assess what they can achieve in light of institutional and budgetary constraints; the experience with political oversight to know what Congress might and might not allow, and what Congress does and does not want them to prioritize; and the best sense of what the universe of regulatory beneficiaries needs from their decisionmaking, not just the few who seek judicial review.¹⁸⁸

Quality assurance initiatives are the epitome of internal administrative law.¹⁸⁹ Following recommendations Mashaw spearheaded for the Administrative Conference of the United States,¹⁹⁰ which the GAO would soon second,¹⁹¹ agencies have used quality assurance programs for decades. Indeed, SSA’s first effort in this regard came shortly after Mashaw argued for internal management strategies as a substitute for individual procedural rights.¹⁹² These experiments offer a tantalizing basis to determine whether internal governance strategies can ensure decisional quality system-wide and thereby succeed where external law falls short.¹⁹³ They also provide an

Intra-Agency Coordination, 129 HARV. L. REV. 421, 426-27 (2015) (providing an example of agency decisionmaking governed by internal administrative law and stressing this law’s importance to agency governance).

185. See Merrill, *supra* note 14, at 57; Metzger & Stack, *supra* note 182, at 1296.

186. See Metzger & Stack, *supra* note 182, at 1263-64.

187. VERMEULE, *supra* note 152, at 115.

188. Cf. *id.* at 124 (“[A]gency procedural decisions are decisions of policy and institutional design, not subject to full legalization.”).

189. See Metzger & Stack, *supra* note 182, at 1253.

190. See generally JERRY L. MASHAW, REPORT IN SUPPORT OF RECOMMENDATION 73-3: QUALITY ASSURANCE SYSTEMS IN THE ADJUDICATION OF CLAIMS OF ENTITLEMENT TO BENEFITS OR COMPENSATION, in 3 RECOMMENDATIONS AND REPORTS OF THE ADMINISTRATIVE CONFERENCE OF THE UNITED STATES: JANUARY 1, 1973-JUNE 30, 1974, at 160.

191. ELMER B. STAATS, U.S. GEN. ACCOUNTING OFFICE, FPCD-78-25, ADMINISTRATIVE LAW PROCESS: BETTER MANAGEMENT IS NEEDED 47 (1978) (recommending the establishment of performance standards for agency adjudication).

192. See Deborah A. Chassman & Howard Rolston, *Social Security Disability Hearings: A Case Study in Quality Assurance and Due Process*, 65 CORNELL L. REV. 801, 801-03, 807-08 (1980).

193. A skeptic might challenge the usefulness of quality assurance initiatives for ALJ-type decisionmaking as a case study for internal administrative law. To Mashaw, hearing-
footnote continued on next page

example of internal administrative law in action, a case study that can help test scholars' claims about internal administrative law's potential and limitations.

III. Quality Assurance in Historical Perspective

We now provide a historical overview of the quality assurance initiatives the three federal agencies have used (or have failed to use, in EOIR's case), based on extensive materials secured through FOIA, in-depth interviews, and administrative records.¹⁹⁴ This previously untold story helps illuminate the factors that have influenced how agencies measure and try to improve decisional quality. The implications of this history come in Part V, but for the moment three observations merit mention. First, SSA's decades of fitful experiences with quality assurance illuminate how concerns not necessarily coterminous with accuracy in decisionmaking, including sensitivity to ALJ independence and overall priorities for policy implementation, determine a program's contours. Second, EOIR's near-total neglect of quality assurance suggests that an agency's commitment to accuracy may depend on the politics of error distribution. If IJ errors tend to disfavor immigrants, the lack of any meaningful quality assurance efforts might reflect the powerlessness of this vulnerable population. Third, BVA's refusal to upgrade its QR program hints

based disability benefits adjudication resulted from a re-creation of agency decisionmaking in judicial guise and imported into SSA a model of justice ill suited to the consistent adjudication of a large number of claims efficiently. He proposed his "internal administrative law" as an alternative. MASHAW, *supra* note 19, at 15. It's no surprise that we have found pathologies in VLJ, ALJ, and IJ governance, one sympathetic to Mashaw's position might maintain, since these sorts of quasi-judicial figures are poorly suited to internal direction. But quasi-judicial decisionmaking for disability benefits, immigration, and veterans' benefits is here to stay. A conception of internal administrative law that declares ALJ-type decisionmaking inapposite risks irrelevance. It also leaves *Mathews v. Eldridge* and individual procedural rights as the sole legal response to crises of decisional quality, an outcome we reject as tantamount to surrender. Finally, BVA and SSA have tried to govern their judges with strategies that fit the internal administrative law mold, suggesting that internal administrative law in action has attempted to accommodate this decisionmaking. *See infra* Parts III.A, .C.

194. Agencies draw from a menu of techniques, including training, peer review, auditing, and internal appeals, among others, to manage the discretion they give frontline adjudicators. *See* Ho & Sherman, *supra* note 24, at 253-65. Agencies have deployed certain of these techniques for programs we do not include within the rubric of quality assurance, even as they bear upon the accuracy of decisionmaking. *E.g.*, Ray & Lubbers, *supra* note 114, at 1599-1601 (describing recent innovations in SSA training techniques). The "quality control system" Mashaw had in mind included "the development of standards, the evaluation of performance against those standards, and action to upgrade substandard performance." Mashaw, *supra* note 17, at 791. We use these characteristics to identify initiatives that attempt to assess the systemic accuracy of agency decisionmaking and improve decisional quality.

at a tension that agencies must resolve as they design their initiatives: A program can paint a rosy picture of accuracy without improving accuracy.¹⁹⁵

A. SSA's Decades of Experiments

SSA has by far the most experience with quality assurance initiatives. Its use of them waxed and waned between the 1970s and mid-2000s. In recent years, the agency has institutionalized its commitment to these programs.

1. Initial efforts

Disability benefits adjudication started in 1956. During the early years, the agency used “own motion review” to examine every allowance decision (a decision by an ALJ granting benefits) and many denials,¹⁹⁶ and it captured systematic data on aspects of ALJ decisionmaking.¹⁹⁷ But SSA lacked any sort of systematic initiative to assess the overall quality of ALJ decisionmaking and recommend improvements.¹⁹⁸ This absence grew glaring as several legal and cultural changes fueled a huge increase in ALJ caseloads in the early 1970s.¹⁹⁹ Despite worrying patterns of ALJ inconsistencies,²⁰⁰ SSA jettisoned its modest quality efforts in 1975 to shift the staff tasked with the work to case production.²⁰¹

195. While we discuss perceptions of the efficacy of various programs that agencies have used throughout this Part, to our knowledge no outside observer has ever subjected any of the initiatives to rigorous and reliable empirical examination, which we turn to in Part IV. While working at BVA, one of us did a superficial internal analysis of its QR program in 2016, reaching similar conclusions about the overall effectiveness of the program to those we make here. This analysis has not been made public, although we quote documents obtained through FOIA that indicate its overall conclusions.

196. See Charles H. Koch, Jr. & David A. Koplow, *The Fourth Bite at the Apple: A Study of the Operation and Utility of the Social Security Administration's Appeals Council*, 17 FLA. ST. U. L. REV. 199, 245 (1990).

197. MASHAW, *supra* note 190, app. at 179-80. Before 1972, SSA ALJs were “hearing examiners.” Ray & Lubbers, *supra* note 114, at 1583.

198. MASHAW, *supra* note 190, app. at 179-80.

199. See *Delays in Social Security Appeals: Hearings Before the Subcomm. on Soc. Sec. of the H. Comm. on Ways & Means*, 94th Cong. 17-18 (1975) [hereinafter *Hearing: Delays*]; DERTHICK, *supra* note 25, at 33-34; Richard E. Levy, *Social Security Disability Determinations: Recommendations for Reform*, 1990 BYU L. REV. 461, 477-478; David E. Rosenbaum, *Huge Federal Disability Program Faces Inequities, Fund Woes, Suits*, N.Y. TIMES (July 27, 1977), <https://perma.cc/TKJ4-32DX>.

200. See Robert G. Dixon, Jr., *The Welfare State and Mass Justice: A Warning from the Social Security Disability Program*, 1972 DUKE L.J. 681, 716-17.

201. *Hearing: Delays*, *supra* note 199, at 75 (statement of James B. Cardwell, Comm'r, Social Security Administration); STAFF OF H. SUBCOMM. ON SOC. SEC., *supra* note 115, at 25. For a description of the pre-1975 process, see STAFF OF H. COMM. ON WAYS & MEANS, 93D CONG., REP. ON THE DISABILITY INSURANCE PROGRAM 136 (Comm. Print 1974).

By 1975, an agency official lamented, “quality control” was “virtually nonexistent.”²⁰²

Amidst this first major “crisis” for the disability benefits adjudication system,²⁰³ the agency announced a management initiative that coupled new case completion goals with a three-pronged quality assurance program.²⁰⁴ Designed to focus on ALJs with aberrational decision patterns, the first two prongs generated instructions tailored to specific ALJs for how they should hold hearings and write decisions.²⁰⁵ The third, dubbed the “Quality Review System,” has become a prototype of sorts and resembles the QR program we examine in-depth in Part IV.²⁰⁶ Beginning in 1976, the agency selected 5% of all ALJ decisions for “post-effectuation” review (i.e., after the agency finalized the decision and notified the claimant).²⁰⁷ Appeals Council analysts looked for errors in these decisions, defined as “any defect so serious that it either required corrective action by the Appeals Council, or would have required such action” had the case been appealed.²⁰⁸ A more searching standard of review, SSA feared, would rankle ALJs.²⁰⁹ A division within SSA used information about “serious” errors analysts observed, along with data culled from other sources, to prepare “quality assurance reports” documenting various trends, which the agency distributed periodically to central and regional supervisory personnel.²¹⁰ The division also used the data to prepare instructional material on specific error areas.²¹¹

202. *Hearing: Delays*, *supra* note 199, at 538 (testimony of Robert L. Trachtenberg, Director, Bureau of Hearings and Appeals).

203. Chassman & Rolston, *supra* note 192, at 807; *see also Hearing: Delays*, *supra* note 199, at 17-18 (describing rising caseload pressures and their impact on decisional quality).

204. STAFF OF H. SUBCOMM. ON SOC. SEC., *supra* note 115, at 18, 27-28; COFER, *supra* note 104, at 84; Jeffrey S. Lubbers, *Federal Administrative Law Judges: A Focus on Our Invisible Judiciary*, 33 ADMIN. L. REV. 109, 125-26 (1981).

205. *See Nash v. Califano*, 613 F.2d 10, 13 (2d Cir. 1980); STAFF OF H. SUBCOMM. ON SOC. SEC., *supra* note 115, at 19-20, 28; COFER, *supra* note 104, at 84-85, 87; Memorandum from Donald A. Gonya, Acting Assoc. Comm’r for Hearings & Appeals, to All Admin. Law Judges 3-4 (July 27, 1979), *reprinted in* Complaint, exhibit F at 25, 27-28, Ass’n of Admin. Law Judges, Judicial Council No. 1 v. Colvin, No. 1:13-cv-02925, 2014 WL 789074 (N.D. Ill. Feb. 26, 2014), ECF No. 1.

206. *See generally* Memorandum from Donald A. Gonya, *supra* note 205, at 2-3 (describing how this program worked).

207. Chassman & Rolston, *supra* note 192, at 813; Memorandum from Donald A. Gonya, *supra* note 205, at 1 (noting that the program began in 1976).

208. Chassman & Rolston, *supra* note 192, at 811.

209. *See id.*

210. *See id.* at 813, 816.

211. *Id.* at 816.

Convinced that SSA cared mostly about case production, ALJs resisted this effort at quality assurance.²¹² They sued to challenge the case completion goals and the prongs that targeted specific ALJs as unlawful interference with their judicial independence.²¹³ The lawsuit settled for an agreement by the agency not to “issue directives or memoranda which set any specific number of dispositions by ALJs as quotas or goals,” and to limit the dissemination of information about individual ALJs and their deficient decisions.²¹⁴ Commenting in their personal capacities, two SSA officials involved with the agency’s quality assurance efforts concluded that the QR System, which the lawsuit did not challenge, achieved little for decisional accuracy.²¹⁵ The initiative did not yield individualized feedback for ALJs, in part due to a concern of a “storm of protest” that might otherwise erupt among them.²¹⁶

2. Bellmon Review

SSA’s second major effort at quality assurance triggered deep and lasting bitterness among ALJs.²¹⁷ Concerned by decisional inconsistencies and the high number of claims ALJs allowed, Congress in 1980 passed the “Bellmon Amendment” to require SSA to restart “own motion” review.²¹⁸ The duty of implementing Bellmon Review fell to the Reagan Administration, which was determined to cut Social Security costs.²¹⁹ Supposedly justified by high error rates in decisions allowing benefits,²²⁰ the initial sampling protocol selected for

212. See STAFF OF H. SUBCOMM. ON SOC. SEC., *supra* note 115, at 24, 32.

213. COFER, *supra* note 104, at 111; DERTHICK, *supra* note 25, at 41.

214. Memorandum from Donald A. Gonya, Acting Assoc. Comm’r for Hearings & Appeals, to All Admin. Law Judges 1 (July 26, 1979), *reprinted in* Complaint, *supra* note 205, exhibit F at 22, 22.

215. See Chassman & Rolston, *supra* note 192, at 817-20.

216. *Id.* at 816; *see also* Mashaw, *supra* note 178, at 823 (“ALJs are too powerful to be subjected to intensive review.”).

217. For evidence of the lingering bitterness, *see, for example, A Proposal to Restructure the Social Security Administration’s Disability Determination Process: Hearing Before the H. Subcomm. on Soc. Sec. of the Comm. on Ways & Means*, 103d Cong. 92-93 (1994) [hereinafter *Hearing: A Proposal*] (statement of David P. Tennant, Treasurer, Association of Administrative Law Judges) (“We remember with horror the pre-effectuation reviews conducted by the Appeals Council in the early 1980s under the name Bellmon review program.”).

218. Social Security Disability Amendments of 1980, Pub. L. No. 96-265, § 304(g), 94 Stat. 441, 456; *Stieberger v. Heckler*, 615 F. Supp. 1315, 1388 (S.D.N.Y. 1985) (noting the name “Bellmon Amendment”), *vacated sub nom. Stieberger v. Bowen*, 801 F.2d 29 (2d Cir. 1986); H.R. REP. NO. 96-944, at 57 (1980) (Conf. Rep.).

219. See DERTHICK, *supra* note 25, at 35-36, 53.

220. See SUBCOMM. ON OVERSIGHT OF GOV’T MGMT. OF THE S. COMM. ON GOVERNMENTAL AFFAIRS, 98TH CONG., REP. ON THE ROLE OF THE ADMINISTRATIVE LAW JUDGE IN THE TITLE II SOCIAL SECURITY DISABILITY INSURANCE PROGRAM 14-15 (Comm. Print 1983).

review 7.5% of ALJ allowance decisions from ALJs who allowed claims most liberally.²²¹ An ALJ's decisions would remain subject to this targeted sampling until his or her accuracy improved, a condition that by clear implication all but required lower allowance rates.²²² SSA claimed to use information gleaned from Bellmon Review to determine training needs and to identify areas where the agency needed to clarify disability policy.²²³ It also proposed to create a feedback system to advise targeted ALJs of "decisional weaknesses"²²⁴ and to put ALJs on improvement plans.²²⁵

Other controversial changes to disability policy caused hearing requests to skyrocket in the early 1980s, creating another caseload crisis for the agency.²²⁶ SSA thus coupled Bellmon Review with a new and much higher case completion goal, increasing each ALJ's target to forty-five decisions per month in 1982.²²⁷ The agency also began to take more muscular action to push ALJs to conform to its expectations for decisionmaking. SSA disclaimed any "improper pressure" to meet allowance or productivity expectations,²²⁸ but it pushed

221. Memorandum from Louis B. Hays, Assoc. Comm'r, Office of Hearings & Appeals, to All Admin. Law Judges 1 (Sept. 24, 1982), reprinted in *Social Security Disability Reviews: The Human Costs; Joint Hearing Before the Subcomm. on Soc. Sec. of the H. Comm. on Ways & Means & the S. Spec. Comm. on Aging*, 98th Cong. 93, 93 (1984) [hereinafter *Hearing: The Human Costs*].

222. SSA insisted that an ALJ with high allowance rates could be removed from review if his decisionmaking proved accurate. *Id.* at 3. But given the agency's conviction that high allowance rates resulted from errors, SSA clearly intended ALJs to lower their allowance rates as a step toward being removed from review. *Cf. Hearing: The Human Costs*, *supra* note 221, at 81 (statement of Jerry Thomasson et al., Administrative Law Judges) (reporting ALJ perceptions that Bellmon Review "targeted" ALJs with high allowance rates).

223. See SUBCOMM. ON OVERSIGHT OF GOV'T MGMT., *supra* note 220, at 15-16.

224. Memorandum from Louis B. Hays, *supra* note 221, at 3.

225. *Id.* at 4.

226. On these changes and their effects, see SUBCOMM. ON OVERSIGHT OF GOV'T MGMT., *supra* note 220, at 2-3; DERTHICK, *supra* note 25, at 33-37; and Levy, *supra* note 199, at 479.

227. Memorandum from Doris A. Coonrod, Reg'l Chief Admin. Law Judge, Region X, to All Admin. Law Judges, Region X, at 1 (July 1, 1982), reprinted in *Social Security Disability Reviews: The Role of the Administrative Law Judge; Hearing Before the Subcomm. on Oversight of Gov't Mgmt. of the S. Comm. on Governmental Affairs*, 98th Cong. 343, 343 (1983) [hereinafter *Hearing: ALJ Role*]. The average ALJ rendered thirty-three decisions in March 1981, then a record for the agency. Memorandum from Andrew J. Young, Assoc. Comm'r, Office of Hearings & Appeals, to Admin. Law Judges (May 27, 1981), reprinted in *Hearing: ALJ Role*, *supra*, at 236.

228. See SUBCOMM. ON OVERSIGHT OF GOV'T MGMT., *supra* note 220, at 1, 16-19.

ALJs informally by allocating support staff and office space, for example, pursuant to performance expectations.²²⁹

By one view, Bellmon Review was not a quality assurance initiative per se, but rather a tool to compromise ALJ independence and pressure ALJs to deny claims.²³⁰ SSA's fixation on case production also crowded out concerns about decisional quality.²³¹ With most ALJs reporting pressure to disallow claims,²³² Bellmon Review prompted another lawsuit on behalf of ALJs complaining of lost decisional independence.²³³ SSA gutted the program, then just starting, before the lawsuit went to trial.²³⁴ The agency replaced its targeted sampling with a protocol that randomly selected cases for review from ALJ decisions allowing and denying claims,²³⁵ and it used the information gleaned from the program to compile bland reports with little guidance to improve decisional quality.²³⁶

Whatever its effect on the accuracy of ALJ decisionmaking,²³⁷ Bellmon Review cast a long, discouraging shadow over quality assurance experiments going forward.²³⁸ Its legacy includes, for instance, regulations amended in 1998

229. *Social Security Restoration Act of 1990: Hearing on S. 2453 Before the Subcomm. on Soc. Sec. and Family Policy of the S. Comm. on Fin.*, 101st Cong. 136 (statement of Eileen P. Sweeney, Staff Attorney, National Senior Citizen Law Center).

230. *See Ass'n of Admin. Law Judges, Inc. v. Heckler*, 594 F. Supp. 1132, 1142 (D.D.C. 1984) (noting that "ALJs could reasonably feel pressure to issue fewer allowance decisions in the name of accuracy"), *amended by* No. 83-0124, 1985 WL 71829 (D.D.C. July 2, 1985).

231. *Cf.* Memorandum from Doris A. Coonrod, *supra* note 227, at 2 ("We should not be required to choose between [quality and quantity] but rather obtain both. However, it is stressed that quantity is necessary to survive.").

232. *Hearing: The Human Costs*, *supra* note 221, at 82-83 (statement of Jerry Thomasson et al., Administrative Law Judges).

233. *Ass'n of Admin. Law Judges*, 594 F. Supp. at 1133.

234. *Stieberger v. Heckler*, 615 F. Supp. 1315, 1379 (S.D.N.Y. 1985), *vacated sub nom.* *Stieberger v. Bowen*, 801 F.2d 29 (2d Cir. 1986).

235. *Id.*; OFFICE OF HEARINGS & APPEALS, REPORT ON QUALITY REVIEW (1986), *reprinted in Administrative Law Judge Corps Act: Hearing on S. 950 Before the Subcomm. on Courts & Admin. Practice of the S. Comm. on the Judiciary*, 100th Cong. 73, 74 (1987).

236. *See, e.g.*, OFFICE OF HEARINGS & APPEALS, *supra* note 235.

237. To be sure, the extent to which Bellmon Review improved accuracy is disputed. *Cf. Stieberger*, 615 F. Supp. at 1379, 1393-94 (suggesting that Bellmon Review may have deprived plaintiffs of unbiased and impartial hearings); U.S. GEN. ACCOUNTING OFFICE, GAO/HEHS-97-102, SOCIAL SECURITY DISABILITY: SSA MUST HOLD ITSELF ACCOUNTABLE FOR CONTINUED IMPROVEMENT IN DECISION-MAKING 19 (1997), <https://perma.cc/9ZUU-E22F> (describing a temporary reduction in ALJ allowance rates but noting that factors other than Bellmon Review might have played a role).

238. *See, e.g.*, *Social Security Disability Programs: Improving the Quality of Benefit Award Decisions; Hearing Before the Permanent Subcomm. on Investigations of the S. Comm. on Homeland Sec. & Governmental Affairs*, 112th Cong. 10 (2012) [hereinafter *Hearing: Decision Quality*] (testimony of Patricia A. Jonas, Deputy Chair, Appeals Council, Social Security Administration).
footnote continued on next page

to disclaim SSA's power to select decisions by individual ALJs or particular hearing offices for pre-effectuation review.²³⁹

3. SSA's modern quality assurance system

SSA used stopgap quality assurance measures from the mid-1980s to 2007, trying but never quite implementing a more expansive program. The agency ended its quality assurance efforts altogether shortly after the Bellmon Review debacle.²⁴⁰ Several times in the 1990s and early- to mid-2000s it planned extensive changes to the disability adjudication process, and in each of these proposals it included revised QR initiatives.²⁴¹ But little materialized, due to inadequate funding and rising backlogs that consumed agency resources.²⁴² One initiative, prompted in part by an explosive 1992 report documenting racial disparities in ALJ allowance rates,²⁴³ amounted to a rebooted random

Security Administration), also available at <https://perma.cc/2XJE-KJG4>; *Hearing: A Proposal*, *supra* note 217, at 92-93 (statement of David P. Tennant, Treasurer, Association of Administrative Law Judges); LEWIN GRP. ET AL., *supra* note 112, at 24. All congressional statements of SSA officials from 2009 to the present are available at *Office of Legislation and Congressional Affairs*, SOC. SECURITY ADMIN., <https://perma.cc/RJ9U-STXN> (archived Oct. 21, 2019).

239. See 20 C.F.R. §§ 404.969, 416.1469 (2019); Administrative Review Process; Identification and Referral of Cases for Quality Review Under the Appeals Council's Authority to Review Cases on Its Own Motion, 63 Fed. Reg. 36,560, 36,570 (July 7, 1998).

240. See *Oversight of the Disability Appeals Process: Hearing Before the Subcomm. on Soc. Sec. of the H. Comm. on Ways & Means*, 105th Cong. 11 (1997) [hereinafter *Hearing: Oversight*] (statement of Carolyn W. Colvin, Deputy Comm'r, Programs and Policy, Social Security Administration) (noting that "[p]rior to 1993, there was no ongoing quality review of hearing decisions per se" and mentioning Appeals Council review of individual cases as the only "quasi-quality review component" in SSA's process).

241. See, e.g., *Social Security Disability Benefits: Did a Group of Judges, Doctors, and Lawyers Abuse Programs for the Country's Most Vulnerable?*; *Hearing Before the S. Comm. on Homeland Sec. & Governmental Affairs*, 113th Cong. 136 (2013) [hereinafter *Hearing: The Most Vulnerable*] (statement of Debra Bice, Chief A.L.J., Social Security Administration), also available at <https://perma.cc/U85U-XPMP> (blaming a lack of quality assurance on the fact that "the Appeals Council was not adequately funded to perform its oversight responsibilities"); *Fourth in a Hearing Series on Securing the Future of the Social Security Disability Insurance Program: Hearing Before the Subcomm. on Soc. Sec. of the H. Comm. on Ways & Means*, 112th Cong. 17-18 (2012) [hereinafter *Hearing: Securing the Future*] (statement of Michael J. Astrue, Comm'r, Social Security Administration), also available at <https://perma.cc/C4PQ-BW4F>; U.S. GEN. ACCOUNTING OFFICE, GAO-02-322, SOCIAL SECURITY DISABILITY: DISAPPOINTING RESULTS FROM SSA'S EFFORTS TO IMPROVE THE DISABILITY CLAIMS PROCESS WARRANT IMMEDIATE ATTENTION 7-9 (2002), <https://perma.cc/2FKL-Y7FC>.

242. See U.S. GEN. ACCOUNTING OFFICE, *supra* note 241, at 25-27.

243. U.S. GEN. ACCOUNTING OFFICE, *supra* note 237, at 20 n.11; see also U.S. GEN. ACCOUNTING OFFICE, GAO/HRD-92-56, SOCIAL SECURITY: RACIAL DIFFERENCE IN DISABILITY DECISIONS WARRANTS FURTHER INVESTIGATION (1992), <https://perma.cc/5W3P-V5V4>.

sampling program along the lines of what the agency tried in 1976.²⁴⁴ Another assigned ALJs to a program of peer review.²⁴⁵ But neither these nor other programs tried during this time grappled with the racial bias issue.²⁴⁶ The random sampling initiative ended in the early 2000s.²⁴⁷ The peer review program concluded in 2009 when ALJs returned to deciding cases.²⁴⁸

In 2006, SSA created an Office of Quality Performance as part of a planned large-scale reconstruction of the disability adjudication process.²⁴⁹ This effort ended in failure a year later, as rising backlogs and congressional pressure forced the agency to shift its priorities.²⁵⁰ It adopted an aggressive backlog reduction plan that included, among other planks, an ALJ hiring spree, a new

244. See *Hearing: Oversight*, *supra* note 240, at 11 (statement of Carolyn W. Colvin, Deputy Comm’r, Programs and Policy, Social Security Administration) (claiming that this program generated “[v]aluable” management information “which has resulted in both ALJ training” and policy reformulation); U.S. GEN. ACCOUNTING OFFICE, *supra* note 237, at 45 & n.b, tbl.4.1; LEWIN GRP. ET AL., *supra* note 112, at 26-27.

245. The agency assigned teams of examiners, medical consultants, and ALJs to scrutinize a random sample of post-effectuation decisions. LEWIN GRP. ET AL., *supra* note 112, at 27. This initiative (1) generated information “to address broad program issues,” *id.*; (2) resulted in the issuance of several reports that indicated the accuracy of ALJ decisionmaking and the reasons for high ALJ award rates, *First in Series on Social Security Disability Programs’ Challenges and Opportunities: Hearing Before the Subcomm. on Soc. Sec. of the H. Comm. on Ways & Means*, 107th Cong. 86 (2001) (statement of Ronald G. Bernoski, A.L.J., Social Security Administration); and (3) ostensibly informed ALJ training, U.S. GEN. ACCOUNTING OFFICE, *supra* note 237, at 20, 64.

246. LINDA G. MILLS, A PENCHANT FOR PREJUDICE: UNRAVELING BIAS IN JUDICIAL DECISION MAKING 59 (1999).

247. LEWIN GRP. ET AL., *supra* note 112, at 26.

248. U.S. GOV’T ACCOUNTABILITY OFFICE, GAO-18-37, SOCIAL SECURITY DISABILITY: ADDITIONAL MEASURES AND EVALUATION NEEDED TO ENHANCE ACCURACY AND CONSISTENCY OF HEARINGS DECISIONS 29 (2017), <https://perma.cc/C7GB-PSHS>.

249. *Social Security’s Improved Disability Determination Process: Hearing Before the Subcomm. on Soc. Sec. of the H. Comm. on Ways & Means*, 109th Cong. 9 (2006) (statement of Jo Anne B. Barnhart, Comm’r, Social Security Administration), also available at <https://perma.cc/R4PX-D62L>. See generally Frank S. Bloch et al., *The Social Security Administration’s New Disability Adjudication Rules: A Significant and Promising Reform*, 92 CORNELL L. REV. 235 (2007) (providing an overview of the redesigned process).

250. See *Hearing: Securing the Future*, *supra* note 241, at 7-8 (statement of Michael J. Astrue, Comm’r, Social Security Administration), also available at <https://perma.cc/C4PQ-BWF4>; U.S. GOV’T ACCOUNTABILITY OFFICE, GAO-08-40, SOCIAL SECURITY DISABILITY: BETTER PLANNING, MANAGEMENT, AND EVALUATION COULD HELP ADDRESS BACKLOGS 43-44 (2007), <https://perma.cc/3JV2-EC8E>; Erik Eckholm, *Disability Cases Last Longer as Backlog Rises*, N.Y. TIMES (Dec. 10, 2007), <https://perma.cc/NW2W-HUHG>.

case completion goal of 500 to 700 decisions per ALJ annually,²⁵¹ and a revived quality assurance program that remains in place today.²⁵²

The first aspect of this program, which a new Division of Quality began in 2010, is the latest iteration of pre-effectuation review—this time of a sample of allowance decisions by a team of Appeals Council analysts.²⁵³ As their predecessors did in the 1970s, staff attorneys gather structured data from each reviewed case on a large array of issues.²⁵⁴ These data help SSA track variances in decision patterns among the country’s various regions and determine which sorts of issues generate the most errors.²⁵⁵ As areas of concern materialize, the Division of Quality can conduct selective samples of decisions, using criteria that can help screen for likely errors.²⁵⁶

SSA uses this pre-effectuation review for several purposes. If it identifies an error in an ALJ’s decision, it may remand that decision to the ALJ for correction.²⁵⁷ Information gleaned from pre-effectuation reviews also helps SSA identify error-prone issues that warrant additional training.²⁵⁸ But the program does not enable SSA to track the performance of individual ALJs or

251. *The Performance of Social Security Administration Appeals Hearing Offices: Hearing Before the Subcomm. on Soc. Sec. of the H. Comm. on Ways & Means*, 110th Cong. 7, 10-12 (2008) (statement of Frank A. Cristaudo, Chief A.L.J., Social Security Administration), also available at <https://perma.cc/NS6U-P5KS>; see also U.S. GOV’T ACCOUNTABILITY OFFICE, GAO-09-398, SOCIAL SECURITY DISABILITY: ADDITIONAL PERFORMANCE MEASURES AND BETTER COST ESTIMATES COULD HELP IMPROVE SSA’S EFFORTS TO ELIMINATE ITS HEARINGS BACKLOG 47-50 (2009), <https://perma.cc/NE63-2V5X> (providing an overview of the backlog reduction plan and its implementation).

252. *Hearing: The Most Vulnerable*, *supra* note 241, at 131-33 (statement of Debra Bice, Chief A.L.J., Social Security Administration), also available at <https://perma.cc/U85U-XPMP>. See generally FELIX F. BAJANDAS & GERALD K. RAY, IMPLEMENTATION AND USE OF ELECTRONIC CASE MANAGEMENT SYSTEMS IN FEDERAL AGENCY ADJUDICATION 42-52 (2018), <https://perma.cc/68BZ-V5RN> (providing an overview of SSA’s quality assurance programs).

253. *Hearing: Decision Quality*, *supra* note 238, at 11 (testimony of Patricia A. Jonas, Deputy Chair, Appeals Council, Social Security Administration), also available at <https://perma.cc/2XJE-KJG4>.

254. *DQ Dives Deep into Quality Data*, EXECUTIVE DIRECTOR’S BROADCAST (Office of Appellate Operations, Soc. Sec. Admin.), Jan. 13, 2012, at 3, reprinted in *Hearing: Decision Quality*, *supra* note 238, at 883, 885.

255. *Id.*

256. OFFICE OF THE INSPECTOR GEN., SOC. SEC. ADMIN., A-12-15-50015, AUDIT REPORT: PRE-EFFECTUATION REVIEWS OF FAVORABLE HEARING DECISIONS 6 n.19 (2017), <https://perma.cc/AGN6-5GHD>.

257. *Id.* at 1.

258. See *id.* at C-2.

hearing offices, avoiding the sort of interference with decisional independence that stymied Bellmon Review.²⁵⁹

The Division of Quality also began “focused quality reviews” in 2011.²⁶⁰ This program is the first since Bellmon Review to scrutinize specific ALJs. It uses an “early monitoring system” with a variety of indicators, including the ALJ’s number of dispositions and the rate at which the Appeals Council agrees with or remands her decisions, to select an ALJ whose decisions indicate a problematic pattern.²⁶¹ If the system is triggered, the Division of Quality may assign a team of attorneys to review a sample of the ALJ’s decisions.²⁶² This review determines if the ALJ’s decisions indeed suffer from a recurring problem, and it helps supervisory personnel identify an appropriate intervention, such as individualized counseling and feedback.²⁶³ From FY 2013 to FY 2016, SSA conducted seventy-two of these time-intensive, focused reviews of ALJs,²⁶⁴ including a number for ALJs with aberrational decision patterns.²⁶⁵ Focused reviews have also helped SSA identify issues that tend to create errors more generally and generate training to address them.²⁶⁶

SSA has developed two more programs in recent years: “disability case reviews” and “inline quality reviews.”²⁶⁷ Its Office of Quality Review has conducted the former by subjecting a representative sample of ALJ decisions to additional scrutiny.²⁶⁸ These post-effectuation reviews, like the pre-effectuation reviews of allowance decisions only, generate information to

259. See *Hearing: The Most Vulnerable*, *supra* note 241, at 131 & n.6 (statement of Debra Bice, Chief A.L.J., Social Security Administration), also available at <https://perma.cc/U85U-XPMP>; see also OFFICE OF THE INSPECTOR GEN., SOC. SEC. ADMIN., A-07-12-21234, CONGRESSIONAL RESPONSE REPORT: THE SOCIAL SECURITY ADMINISTRATION’S REVIEW OF ADMINISTRATIVE LAW JUDGES’ DECISIONS 8 (2012), <https://perma.cc/LS2B-FHF7> (describing examples of what the program, as of 2012, doesn’t track).

260. *Hearing: The Most Vulnerable*, *supra* note 241, at 132 (statement of Debra Bice, Chief A.L.J., Social Security Administration), also available at <https://perma.cc/U85U-XPMP>.

261. U.S. GOV’T ACCOUNTABILITY OFFICE, *supra* note 248, at 31.

262. *Id.*; see also OFFICE OF THE INSPECTOR GEN., SOC. SEC. ADMIN., A-12-16-50106, AUDIT REPORT: OVERSIGHT OF ADMINISTRATIVE LAW JUDGE DECISIONAL QUALITY, at D-2 (2017), <https://perma.cc/42WF-K6VE>.

263. OFFICE OF THE INSPECTOR GEN., *supra* note 262, at D-2; see Ray & Lubbers, *supra* note 114, at 1598-99.

264. OFFICE OF THE INSPECTOR GEN., *supra* note 262, at D-2 n.8.

265. *Id.* at D-2; “*Focused Quality Reviews*” Enhance DQ’s Mission, EXECUTIVE DIRECTOR’S BROADCAST (Office of Appellate Operations, Soc. Sec. Admin.), Jan. 13, 2012, at 1, reprinted in *Hearing: Decision Quality*, *supra* note 238, at 883, 886.

266. See OFFICE OF THE INSPECTOR GEN., SOC. SEC. ADMIN., A-12-15-15005, THE SOCIAL SECURITY ADMINISTRATION’S EFFORTS TO ELIMINATE THE HEARINGS BACKLOG, at B-2 to -3 (2015), <https://perma.cc/AA9P-GX7J>.

267. U.S. GOV’T ACCOUNTABILITY OFFICE, *supra* note 248, at 35.

268. *Id.* at 37 tbl.2.

guide ALJ training.²⁶⁹ Pursuant to the inline QR program, begun in 2009, staff in SSA's ten regional offices randomly sample cases in hearing offices at various stages before effectuation.²⁷⁰ These reviews examine the extent to which hearing office personnel, including staff and decision writers in addition to ALJs, prepare cases in a manner consistent with SSA policy.²⁷¹

Recognizing the proliferation of quality assurance initiatives and the need to coordinate them, SSA created the Office of Analytics, Review, and Oversight in October 2017.²⁷² As this institutionalization of quality assurance suggests, SSA's post-2007 efforts are the most enduring since the agency began its experiments in the late 1970s. SSA, for instance, has also developed novel uses of data analytics and natural language processing to improve quality and case processing times.²⁷³ Some signs suggest an impact on ALJ decision patterns,²⁷⁴ but an external evaluation of its quality effects has not been conducted.²⁷⁵

B. EOIR's Inattention

While SSA's experimentation with quality assurance initiatives has been extensive, EOIR's has remained superficial. Its near-total disregard for quality assurance initiatives may reflect immigrants' political weakness and may thereby shed as much light on the pressures that affect internal regimes of administrative governance as the decades of SSA experiments do.

269. *See id.* at 39.

270. *Id.* at 37 tbl.2.

271. *Id.* at 32, 35-36.

272. *Determining Eligibility for Disability Benefits: Challenges Facing the Social Security Administration; Hearing Before the Subcomm. on Soc. Sec. of the H. Comm. on Ways & Means, 115th Cong. 20-21 (2017)* (statement of Bea Dismar, Acting Chief of Staff, Social Security Administration), also available at <https://perma.cc/237T-424D>.

273. *See* BAJANDAS & RAY, *supra* note 252, at 49-51; Ray & Lubbers, *supra* note 114, at 1591. For a description of the early stages of SSA's analytics efforts, see Verkuil, *supra* note 27, at 244-45.

274. *See* Ray & Lubbers, *supra* note 114, at 1604-07. Other developments, including a rapid expansion of the ALJ corps and significant changes to ALJ training, complicate an assessment of exactly how much credit the quality assurance initiatives deserve. *See* U.S. GOV'T ACCOUNTABILITY OFFICE, *supra* note 248, at 25; 2018 SOC. SECURITY ADMIN., *supra* note 85, at 2.79 tbl.2.F8.

275. *See* OFFICE OF THE INSPECTOR GEN., SOC. SEC. ADMIN., A-12-18-50353, AUDIT REPORT: THE SOCIAL SECURITY ADMINISTRATION'S USE OF INSIGHT SOFTWARE TO IDENTIFY POTENTIAL ANOMALIES IN HEARING DECISIONS 11 (2019), <https://perma.cc/2BH5-7P25> (concluding that "SSA was not regularly tracking management information to determine whether Insight was meeting its goals").

The earliest program appears to be EOIR's "Immigration Court Evaluation Program."²⁷⁶ The agency sent a team of IJs, legal assistants, and court administrators to various immigration courts for qualitative peer evaluation.²⁷⁷ This program ended in 2008, when EOIR replaced the review process with a set of quantitative measures of court performance that focus on case completion goals.²⁷⁸ Performance measures adopted in 2018 for IJs focus almost entirely on case production and wholly neglect decisional accuracy.²⁷⁹ EOIR appears to rely upon Assistant Chief Immigration Judges, who supervise regions of IJs, to determine which IJs need some intervention to improve performance.²⁸⁰ It provides no guidance, however, to these supervisory judges as to how they should evaluate IJ performance, and thus no guarantee evaluations proceed uniformly.²⁸¹

One episode illustrates just how little attention EOIR has paid to quality assurance. In 2002, the U.S. Attorney General implemented "streamlining" at

276. In 2008, a GAO report identified IJ training and the one-by-one review of IJ decisions by the Board of Immigration Appeals as EOIR's only centrally administered programs with decisional quality as a primary concern. See U.S. GOV'T ACCOUNTABILITY OFFICE, GAO-08-935, U.S. ASYLUM SYSTEM: AGENCIES HAVE TAKEN ACTIONS TO HELP ENSURE QUALITY IN THE ASYLUM ADJUDICATION PROCESS, BUT CHALLENGES REMAIN 19 (2008), <https://perma.cc/6X4A-KUCZ>. We submitted a FOIA request in January 2016 asking for "[a]ll studies, memoranda, reports, or other documents produced from January 1, 2010, to the present discussing, describing, analyzing, or proposing agencywide quality assurance initiatives offered or designed to improve the accuracy and timeliness of immigration judge decision-making." Letter from David Marcus, Professor of Law, Univ. of Ariz., to Office of the Gen. Counsel, Exec. Office for Immigration Review 2 (Jan. 28, 2016) (on file with authors). We appealed when EOIR's production in response to this request indicated no more than what we summarize here, but we did not receive any additional information. Letter from David Marcus, Professor of Law, Univ. of Ariz., to Dir., Office of Info. Policy (Mar. 2, 2017) (on file with authors). We also attempted to interview EOIR personnel but were not permitted to do so. We are thus reasonably confident that our summary accurately reflects the history of EOIR quality assurance, although it is possible that EOIR materials from before January 1, 2010 that are not reflected in publicly available sources memorialize more extensive efforts.

277. U.S. GOV'T ACCOUNTABILITY OFFICE, GAO-17-438, IMMIGRATION COURTS: ACTIONS NEEDED TO REDUCE CASE BACKLOG AND ADDRESS LONG-STANDING MANAGEMENT AND OPERATIONAL CHALLENGES 60 (2017), <https://perma.cc/N4BB-3DAZ>.

278. *Id.* at 60-61. In 2017, EOIR insisted that a new "Organizational Results Unit" that supposedly would have started at the end of 2017, would operate as a successor to this program. *Id.* at 66. We are unable to determine whether or not this unit launched as planned.

279. See Memorandum from James R. McHenry III, Dir., Exec. Office for Immigration Review, to Office of the Chief Immigration Judge et al. app. A (Jan. 17, 2018) (on file with authors).

280. U.S. GOV'T ACCOUNTABILITY OFFICE, GAO-08-940, U.S. ASYLUM SYSTEM: SIGNIFICANT VARIATION EXISTED IN ASYLUM OUTCOMES ACROSS IMMIGRATION COURTS AND JUDGES 9 (2008), <https://perma.cc/C7BW-5F3C>.

281. See *id.*

the Board of Immigration Appeals (BIA), the internal appellate body that reviews IJ decisions.²⁸² To eliminate BIA backlogs (among other objectives), the Attorney General authorized single BIA members, instead of panels, to decide appeals much more often, and with much less reasoning.²⁸³ Shoddy agency review prompted immigrants to flood the federal courts with appeals.²⁸⁴ Federal judges repeatedly lashed out at poor-quality decisionmaking within EOIR, a backlash that pushed the Attorney General to respond.²⁸⁵ In August 2006, he ordered EOIR to make twenty-two changes to improve immigration adjudication.²⁸⁶ These included, among others, the adoption of “mechanisms to detect poor conduct and quality.”²⁸⁷

A few months later, EOIR’s Director issued a memorandum detailing the status of each required change. He noted that EOIR had “[i]mplemented” the quality assurance directive by issuing an instruction that the Department of Justice lawyers who defend immigration orders in immigration court report instances when “an immigration judge failed to display the appropriate level of professionalism.”²⁸⁸ Even in response to a crisis of legitimacy that threatened to unmake core aspects of the immigration court system,²⁸⁹ one prompted by disastrous decisionmaking quality in the immigration courts, EOIR declined to fashion a quality assurance program of the sort that SSA had used since the 1970s.

C. BVA and Quality Review

Along the spectrum of experience and innovation, BVA fits between SSA and EOIR. Its institutional commitment to quality assurance goes back decades,²⁹⁰ but its chief program, which took shape in 1998, is basic in

282. See Katie R. Eyer, *Administrative Adjudication and the Rule of Law*, 60 ADMIN. L. REV. 647, 673-74 (2008) (describing the “streamlining” episode).

283. *Id.*

284. See Bert I. Huang, *Lightened Scrutiny*, 124 HARV. L. REV. 1109, 1122-23 (2011).

285. *EOIR Revises Streamlining Fact Sheet*, 83 INTERPRETER RELEASES 512, 512 (2006).

286. Memorandum from Alberto Gonzales, Attorney Gen., to the Deputy Attorney Gen. et al. (Aug. 9, 2006) (on file with authors).

287. *Id.* at 3 (capitalization altered).

288. Memorandum from Kevin D. Rooney, Dir., Exec. Office for Immigration Review, to All Exec. Office for Immigration Review Emps. 3 (Mar. 2007), <https://perma.cc/JCS7-VEDB>.

289. See generally MARGARET MIKYUNG LEE, CONG. RESEARCH SERV., RL33410, IMMIGRATION LITIGATION REFORM (2006), <https://perma.cc/5S3S-BDVF> (discussing proposed changes to the review of IJ decisions prompted in part by the streamlining experience).

290. See Memorandum from Charles L. Cragin, Chairman, Bd. of Veterans’ Appeals, to Dir., Compensation & Pension Serv., Bd. of Veterans’ Appeals et al. 1 (May 9, 1996) (on file with authors).

design.²⁹¹ The program has generated a blunt accuracy number that invariably casts BVA performance in a favorable light—a feature that perhaps explains why the program has proven resistant to change.²⁹²

Prior to CAVC's creation in 1988, the "splendid isolation"²⁹³ of the veterans' benefits system made external review of quality difficult.²⁹⁴ CAVC added significantly more scrutiny to BVA's work and pressed it to ensure that VLJs rendered policy-compliant decisions.²⁹⁵ This additional scrutiny had an immediate and dramatic impact on BVA's substantive work, resulting in a significant drop in the denial rate and an increase in the rate at which VLJs remanded cases.²⁹⁶ As BVA's chairman acknowledged in 1996, CAVC "has provided the most significant guidance in defining the essential ingredients of quality appellate decisions."²⁹⁷

Before 1996, BVA used a quality system that "subjected all Board decisions to screening for errors before promulgation and approximately 10 to 15 percent of those decisions to a more intense quality review."²⁹⁸ But it failed to keep statistics or data on the results of these reviews,²⁹⁹ and thus, by BVA's own admission, this system "did not lend itself to quantifiable measurement and repeatable comparisons over time."³⁰⁰ BVA revised the system in 1998 to create a quantitative program to evaluate and score BVA's decisionmaking accuracy, and to collect data to identify areas needing improvement.³⁰¹

291. A number of the institutional details documented in this Subpart are based on personal knowledge and experience by one of us, who served as Chief of the Office of Quality Review for BVA. Not all of these details have publicly available citations.

292. *See infra* Part IV.

293. *Brown v. Gardner*, 513 U.S. 115, 122 (1994) (quoting H.R. REP. NO. 100-963, pt. 1, at 10 (1988)); *see also* Robert L. Rabin, *Preclusion of Judicial Review in the Processing of Claims for Veterans' Benefits: A Preliminary Analysis*, 27 STAN. L. REV. 905, 905 (1975) (coining the term "splendid isolation").

294. For an overview of quality reporting in the years before CAVC's creation, *see* Ridgway, *supra* note 77, at 255-56.

295. 1992 BOARD VETERANS' APPEALS, *supra* note 80, at 10.

296. *See* Ridgway & Ames, *supra* note 149, at 325.

297. Memorandum from Charles L. Cragin, *supra* note 290, at 2.

298. *Id.* at 1.

299. Richard Thrasher, Chief Counsel for Policy, Bd. of Veterans' Appeals, *The BVA: Who We Are and Why We're Here*, Seventh Judicial Conference: United States Court of Appeals for Veterans Claims (Sept. 19-20, 2002), in 17 VETERANS APPEALS REP., at LXXXI, CXI (2004).

300. 1998 BOARD VETERANS' APPEALS REP. CHAIRMAN 25.

301. *Id.* *See generally* Memorandum from Richard B. Standefer, Acting Chairman, Bd. of Veterans' Appeals, to Dir., Compensation & Pension Serv., Bd. of Veterans' Appeals et al. 1 (May 14, 1998) (on file with authors) (describing the program's workings). The stated purposes of the QR system were to (1) "measure performance in the area of quality for the Board as a whole and for each of the four decision teams;" and (2) "collect
footnote continued on next page

The QR program that took form in 1998 ran continuously until November 2016.³⁰² It substantially resembled the “Quality Review System” SSA debuted in 1976.³⁰³ In 2002, a BVA computer system was set to randomly select for review 5% of all original BVA decisions and 10% of cases remanded by CAVC.³⁰⁴ The system would make these selections after the VLJ had completed the draft decision, but before issuing it to the appellant. Prior to May 2002, staff who remained involved in deciding cases reviewed decisions selected by the QR program.³⁰⁵ The GAO raised concerns about this structure in 2002, suggesting that BVA members directly involved in deciding veterans’ appeals should not also review draft decisions for accuracy.³⁰⁶ BVA revised the program in May 2002, concentrating all QR work in its Appellate Group,³⁰⁷ a managerial body not typically involved in direct adjudication.³⁰⁸ Within the group, the Office of Quality Review,³⁰⁹ a team of four to six competitively selected attorneys serving two-year details, administered the program.

data to identify areas where quality of decisionmaking needed improvement . . . [.] set realistic goals for improvement, and . . . develop training programs.” Thrasher, *supra* note 299, at CXII. From 1995 to 2012, VLJs were divided into four “Decision Teams,” each managed by one Deputy Vice Chairman and two Chief VLJs. 2013 BOARD VETERANS’ APPEALS ANN. REP. 3. These teams divided workload along geographical lines and each Decision Team took all appeals from specific VA regional offices. 1995 BOARD VETERANS’ APPEALS, *supra* note 83, at 9. In December 2012, the Decision Teams were abolished and the VLJs were reorganized into a single body headed by two Deputy Vice Chairmen and ten Chief VLJs. 2013 BOARD VETERANS’ APPEALS, *supra*, at 3.

302. See Email from Carol DiBattiste, to BVA List & BVA Wilkes-Barre (Nov. 16, 2016, 4:56:53 PM) (on file with authors) (announcing new QR program to BVA employees).

303. See *supra* text accompanying notes 206-08.

304. Email from Richard C. Thrasher, Chief Counsel for Policy, Bd. of Veterans’ Appeals, to BVA List (May 15, 2002, 3:49 PM) (on file with authors); see also CYNTHIA A. BASCETTA, U.S. GOV’T. ACCOUNTABILITY OFFICE, GAO-05-655T, VA DISABILITY BENEFITS: BOARD OF VETERANS’ APPEALS HAS MADE IMPROVEMENTS IN QUALITY ASSURANCE, BUT CHALLENGES REMAIN FOR VA IN ASSURING CONSISTENCY 7 (2005), <https://perma.cc/R2B3-5KR7>. Prior to May 2002, 100% of cases which had previously been remanded by CAVC were selected for QR review. See Email from Richard C. Thrasher, *supra*.

305. See U.S. GEN. ACCOUNTING OFFICE, GAO-02-806, VETERANS’ BENEFITS: QUALITY ASSURANCE FOR DISABILITY CLAIMS AND APPEALS PROCESSING CAN BE FURTHER IMPROVED 7 n.7 (2002), <https://perma.cc/WBL7-HBTZ>.

306. *Id.*; see also BASCETTA, *supra* note 304, at 7 (describing this history).

307. BVA’s Appellate Group was renamed the Office of the Principal Deputy Vice Chairman in FY 2017. 2017 BOARD VETERANS’ APPEALS, *supra* note 1, at 8; 2016 BOARD VETERANS’ APPEALS, *supra* note 23, at 3-4.

308. See 2002 BOARD VETERANS’ APPEALS REP. CHAIRMAN 6-7; U.S. GEN. ACCOUNTING OFFICE, *supra* note 305, at 7 n.7.

309. At various times, the Office of Quality Review was also referred to as the Quality Review Unit and the Office of Quality Assurance. For consistency, this Article will only refer to it as the Office of Quality Review.

Between 1998 and 2002, a case was considered erroneous, or inaccurate, if it contained at least one error from a list of predefined categories, internally known as “exception[s].”³¹⁰ The error standard explicitly omitted any situation in which there could be “[l]egitimate differences of opinion as to the outcome in an appeal, the interpretation of the law, the application of the law to the facts, or assessment of the weight and credibility of the evidence.”³¹¹ As a result of the 2002 GAO report, BVA narrowed the definition of an error to “a deficiency that would be outcome determinative, that is, result in the reversal or remand of a Board decision by the [CAVC].”³¹²

Between 2003 and 2016, BVA’s accuracy rate never dipped below 89%, even as the rate at which CAVC remanded or reversed BVA decisions never fell below 44%.³¹³ This apparent disparity prompted skepticism about the accuracy rate’s integrity.³¹⁴ In 2016, one of us, who headed the QR program at the time, proposed replacing the existing system with a data-driven system to identify recurring problems and generate interventions to address these problems.³¹⁵ As the proposal indicated, “The accuracy rate number [the existing program generates] . . . is essentially static, does not reflect any long-term trends or changes in quality, and is completely inconsistent with results reported by the

310. See 2002 BOARD VETERANS’ APPEALS, *supra* note 308, at 7. For a list of these error categories, see U.S. GEN. ACCOUNTING OFFICE, *supra* note 305, at 8, 9 tbl.2. In August 2015, after overhauling its coding methodology, BVA reduced the six categories to five: (1) issues, (2) remands, (3) reasons or bases, (4) due process, and (5) miscellaneous. U.S. DEP’T OF VETERANS AFFAIRS, VETERANS APPEALS CONTROL AND LOCATOR SYSTEM (VACOLS) 3 DATA DICTIONARY 49 (2018) (on file with authors).

311. Memorandum from Richard B. Standefer, *supra* note 301, at 3.

312. 2002 BOARD VETERANS’ APPEALS, *supra* note 308, at 7. This change eliminated all errors in the “Format” category. See BASCETTA, *supra* note 304, at 7-8.

313. BVA reports its accuracy rates in annual reports, and CAVC reports remand and reversal rates in its annual reports. See *The Board of Veterans Appeals Annual Reports to Congress: 1991-2017*, U.S. DEP’T VETERANS AFF., <https://perma.cc/FR6F-82ZR> (last updated Jan. 8, 2019); *Court Reports*, U.S. CT. APPEALS FOR VETERANS CLAIMS, <https://perma.cc/GYL5-3A7M> (archived Oct. 26, 2019); see also 2 U.S. DEP’T OF VETERANS AFFAIRS, FY 2008 BUDGET SUBMISSION, at 7C-2 tbl. (2007), <https://perma.cc/2U7L-SYB4> (providing accuracy rates for 2003 to 2006).

314. See *Board of Veterans’ Appeals Adjudication Process and the Appeals Management Center: Hearing Before the Subcomm. on Disability Assistance & Mem’l Affairs of the H. Comm. on Veterans’ Affairs*, 110th Cong. 2 (2007) [hereinafter *Hearing: BVA Adjudication*] (statement of Rep. John J. Hall, Chairman) (“[A]lthough the BVA claims a 93 percent accuracy rate, the U.S. Court of Appeals for Veterans Claims (CAVC) sets aside or remands over 70 percent of the cases appealed indicating a much lower accuracy rate in reality.”); Memorandum from Steven L. Keller, Vice Chairman, Bd. of Veterans’ Appeals, to Deputy Gen. Counsel 1 (Aug. 3, 2010) (on file with authors) (describing VA’s Office of General Counsel’s skepticism of the accuracy rate in light of a high CAVC remand rate).

315. Office of Quality Assurance, Bd. of Veterans’ Appeals, QA Metrics Proposal (n.d.) (on file with authors).

CAVC.”³¹⁶ BVA experimented with this alternative system for just enough time to generate an in-depth study of a particular regulation that had proven difficult for VLJs to apply properly.³¹⁷ Subsequent training based on this study suggested that VLJs change their behavior in this area of law.³¹⁸

In late 2017, however, BVA reinstated a version of its old QR program.³¹⁹ As one official told us, a single erroneous VLJ decision had embarrassed VA, and in response BVA’s Interim Principal Deputy Vice Chairman³²⁰ insisted that it restart the 2002 version of the QR program.³²¹ BVA’s reconstituted program samples a smaller percentage of draft decisions, as increasing caseloads made sampling 5% of original decisions difficult absent an increase in resources.³²² Previously, attorneys assigned to the program served two-year details, as an attorney “took 6-12 months to gain expertise.”³²³ The reinstated program involved attorneys on six-month details, with the “[g]oal to achieve full productivity within *one week* of training.”³²⁴

Importantly, the reconstituted program uses a new standard of review that abandons the pretense that reviewing attorneys predict CAVC remands.

316. *Id.*

317. See generally OFFICE OF QUALITY ASSURANCE, BD. OF VETERANS’ APPEALS, TQA 17-1, TARGETED QUALITY ANALYSIS: APPEALS OF EXTRASCHEDULAR RATINGS UNDER 38 C.F.R. § 3.321(b)(1) (2017) (on file with authors).

318. See Office of Quality Assurance, Bd. of Veterans’ Appeals, Navigating § 3.321(b)(1) Extraschedular Analysis (Aug. 2017) (presentation on file with authors). We are able to observe a drop in the citations to the relevant case and regulation after the training by examining all BVA decisions before and after the training. This drop suggests that the instruction in the training to cite the case and regulation more sparingly was effective.

319. See 2018 BOARD VETERANS’ APPEALS, *supra* note 72, at 11 (describing the “launch of a new iteration of the Board’s accuracy rate measurement system”); see also Office of Quality Assurance, Bd. of Veterans’ Appeals, Case Review Modernization 2, 4-7 (Sept. 2017) (presentation on file with authors) (indicating that core features, such as staffing and random sampling of pre-issuance decisions to correct errors remained the same, despite changes in the term of staff attorneys and a formal change in the error standard).

320. The Interim Principal Deputy Vice Chairman, Cheryl L. Mason, was subsequently nominated by President Trump as the Chairman and confirmed in November 2017. See *Cheryl L. Mason*, U.S. DEP’T VETERANS’ AFF., <https://perma.cc/8JPF-63U4> (archived Oct. 26, 2019).

321. Telephone Interview with Interview Subject No. 7 (Sept. 28, 2018). There is no obvious reason why jettisoning the old system produced the erroneous VLJ decision, and thus why the erroneous VLJ decision would justify readopting the old system. The only conceivable reason why BVA responded in this way was that it wanted to be seen as taking decisive action with respect to quality.

322. See 2018 BOARD VETERANS’ APPEALS, *supra* note 72, at 11; Office of Quality Assurance, *supra* note 319, at 2.

323. Office of Quality Assurance, *supra* note 319, at 6.

324. *Id.* at 7.

Instead, the new standard requires that errors be “clear and unmistakable”—ones where “the result would have been manifestly different but for the error.”³²⁵ Such instances should be “very, very rare.”³²⁶ Using this new approach, BVA reported a 93.6% accuracy rate for FY 2018.³²⁷ During these twelve months, a remarkable 47% increase in staff productivity enabled BVA to decide 85,288 cases, crushing its record for FY 2017.³²⁸

* * *

We can draw several lessons from this untold story of quality management in the administrative state. First, the variability in agency attention to quality is substantial. Despite the constitutional demand for accuracy and evidence of considerable challenges at each of these agencies, the variation is striking. As SSA’s Office of Analytics, Review, and Oversight demonstrates, quality assurance has become institutionalized at SSA after decades of fitful experiments. To this day, BVA continues to use a model rooted in the 1970s, notwithstanding an internal push for more data-driven techniques. EOIR equates unprofessional conduct by IJs with low quality, suggesting a hollowing out of *Goldberg’s* demand for accuracy within the agency. These varying quality efforts may be indicative of differences in motivation. SSA’s numerous quality program changes and experiments suggest an honest search for accurate quality measurement and improvement by at least some portion of the agency. BVA’s utilization of the same program, without any notable changes, for over fifteen years suggests that the rosy statistics the program produces are the true goal. EOIR’s near-total lack of quality-monitoring efforts reflects an agency unconcerned with accountability for poor quality.

Second, quality remains a deeply contested concept. In the 1980s, SSA deployed it for political control to reduce the welfare rolls, and quality assurance initiatives continue to challenge core notions of decisional independence. BVA’s reforms illustrate the challenges of internal institutional design around quality, with GAO critiquing the use of staff attorneys to review decisions by superiors. As we show in more depth in Part V, bureaucratic incentives can distort how an agency measures and reports quality. While few would contest that agencies should strive for high-quality decisions at an abstract level, the implementation is anything but straightforward.

325. *Id.* at 8; Office of Quality Assurance, Bd. of Veterans’ Appeals, QR Error Standards—Summary Only (n.d.) (on file with authors) (capitalization altered).

326. Office of Quality Assurance, *supra* note 319, at 9.

327. 2018 BOARD VETERANS’ APPEALS, *supra* note 72, at 11.

328. *Id.* at 8-9, 9 fig.; *id.* at 10-11 (describing staffing and caseload increase); *see also* Ridgway, *supra* note 2, at 8 (charting BVA productivity from 1997 to 2018).

Third, while there has been much experimentation, we know virtually nothing about which initiatives actually work. This is unfortunate as it undercuts the ability of all federal agencies to learn how to design effective adjudicatory systems. SSA insiders suggest significant improvement at SSA's Appeals Council, but the effectiveness of interventions for ALJ decisionmaking remains uncertain.³²⁹ No quality assurance program has even been subjected to an independent evaluation. Yet if institutional and doctrinal limitations hobble the capacity of external law to ensure an adequate threshold of decisional quality in mass adjudication, and if internal administrative law offers the best alternative, it is critical to understand whether these initiatives are as effective as billed.

IV. Empirical Case Study: BVA's Quality Review Program

As one of us previously served as BVA's Director of the Office of Quality Review, we are able to draw on this institutional insight to offer, for the first time, a rigorous assessment of whether this form of internal administrative law works. In this Part, we first spell out what makes our case study unique and then highlight the main findings. Our takeaway is straightforward: BVA's QR program failed. Cases that went through QR were no less likely to be appealed and remanded than cases that did not. The likely reason for this stunning finding involves the standard of review QR attorneys used when they evaluated VLJ decisions. While ostensibly predicting what would happen to cases if appealed to CAVC, these attorneys in reality were a good deal more deferential to VLJs. The QR program generated rosy numbers, suggesting highly accurate decisionmaking that bore little relationship to any external measure of quality. It gave VLJs virtually no incentive to improve.

A. A Critical Test Case

Our case study is unique in several respects. First, as spelled out in Part III, BVA's QR program remains the leading prototype of how to assess and improve the accuracy of decisionmaking through an internal agency program. Second, our examination is the first independent test of such a program.³³⁰ Third, in contrast to previous studies that are limited to reporting highly aggregated data from agency output, we secure rich, internal data of every case that the agency adjudicated from 2002 to 2016, comprising nearly 600,000 cases.

329. See BAJANDAS & RAY, *supra* note 252, at 47-48.

330. We provide the technical aspects and more extensive statistical results in our companion article for a social science audience. See generally Daniel E. Ho et al., *Quality Review of Mass Adjudication: A Randomized Natural Experiment at the Board of Veterans Appeals, 2003-16*, 35 J.L. ECON. & ORG. 239 (2019).

As discussed in Part III, the QR program ran formally unchanged from November 1, 2002, to November 15, 2016, giving us a long observation period to test for subsequent disposition of cases.³³¹ The “Veterans Appeals Control and Locator System” (VACOLS) dataset includes rich information about every veteran’s case—including service period, fine-grained medical diagnosis, and the subsequent disposition of the case at CAVC if appealed.³³² It is the same dataset that BVA used to administer the program, allowing us to construct rigorous statistical tests for its efficacy. These rich data mean that we are able to provide very precise estimates of effect sizes. Fourth, the design of the program allows us to construct a highly credible test for its effectiveness. Most importantly, a computer *randomly* selected 5% of original cases (VLJ decisions in cases before BVA for the first time) for QR. Random assignment ensures that those cases selected for QR are comparable in all respects to those not selected for QR, allowing us to attribute outcome differences to the QR intervention. Fifth, our data indicate which cases were selected for QR and what errors, if any, the QR team identified for selected cases. VACOLS thus enables us to divide all BVA decisions into a control group (cases not selected for QR) and a treatment group (cases selected for QR).

Random assignment, coupled with these rich data, hence allows us to test whether the program worked as intended, by comparing QR cases and control cases (i.e., those not subjected to QR). If the program worked, we would expect veterans to appeal a much smaller percentage of QR cases than control cases. To determine whether to “call” a “substantive error”—basically a problem other than a formatting mistake—the QR attorney reviewing the draft decision would decide whether the opinion exhibited “a deficiency that would be outcome determinative, that is, result in the reversal or remand of a [BVA] decision by the [CAVC].”³³³ If she found such an error, the QR team would prepare and circulate an “exception memorandum” describing the error for the VLJ.³³⁴ The VLJ could revise the draft decision to correct the error, ignore the exception memorandum, or make a formal challenge to BVA’s Chief Counsel for Policy and Procedure.³³⁵ The express purpose was to correct errors so as to reduce the likelihood of an appeal to, and reversal or remand by, CAVC.³³⁶ In addition, we would expect CAVC to reverse or remand a smaller number of

331. *See supra* notes 302-18 and accompanying text.

332. U.S. DEP’T OF VETERANS AFFAIRS, *supra* note 310, at 3, 8, 18.

333. 2002 BOARD VETERANS’ APPEALS, *supra* note 308, at 7; *see supra* note 312 and accompanying text.

334. Paul Sorisio, Focus on Quality: The Internal Workings of Quality Review (QR), Remarks at National Veterans Service Advanced Skill Level Training 7, 9 (Dec. 4, 2011) (presentation on file with authors).

335. *Id.* at 11.

336. *See id.* at 7; *see also* 2002 BOARD VETERANS’ APPEALS, *supra* note 308, at 7.

issues that underwent QR. Exception memoranda would have prompted corrections in the small number of decisions with errors before they went out the door.

B. Results and Analysis

Our basic findings are damning: QR had no causal effects for either original appeals or CAVC-remanded appeals. Veterans appealed 6% of VLJ decisions on original cases in the treatment group and 6% of the cases from the control group to CAVC.³³⁷ Again, due to the size of the sample, these estimates are precise, allowing us to rule out the QR program having effects of any substantial magnitude. The top panel of Figure 1 plots the proportion of QR and control cases appealed to CAVC in red and blue, respectively, with 95% confidence intervals that overlap. In addition, our tests show that QR had virtually no impact on the likelihood that CAVC would reverse or remand a BVA decision. Of original cases appealed to CAVC, roughly 75% of QR cases were vacated and remanded by CAVC, compared to 76% of control cases. The bottom panel of Figure 1 displays the set of typical outcomes for QR and control cases, in red and blue respectively, showing no differences of any substantive magnitude.

337. Through a series of robustness analyses, we find marginal evidence that the QR program may have had an effect for cases back on remand from CAVC. QR may have caused a 1% reduction in the rate at which veterans took decisions in CAVC-remanded cases back to CAVC again. *See* Ho et al., *supra* note 330, at 253-54. With 5,622 CAVC-remanded cases undergoing QR, we estimate that the program has helped BVA avoid roughly sixty appeals over fifteen years. *Id.* at 254. This effect is de minimis: During this time, veterans appealed 57,170 cases to CAVC in total. These figures are in CAVC's Annual Reports, 2003 to 2017, which are available on CAVC's website. *See Court Reports, supra* note 313.

Figure 1
The Impact of QR on Case Outcomes at CAVC

In the top panel, we plot the rate at which cases that went through QR (in red) and control cases (in blue) are appealed to CAVC. In the bottom panel, we plot the proportion of CAVC decisions in which at least one issue was subject to each disposition (e.g., vacated and remanded), again separately for cases that went through QR (in red) and control cases (in blue). Vertical dashes indicate the point estimates and horizontal lines indicate the 95% confidence interval. These results show that the QR intervention had no substantive or statistically significant effect on CAVC outcomes: The QR and control outcomes are all comparable. Because the sample size involves 26,821 QR cases and 508,801 control cases, our estimates are extremely precise.

Why did the QR program fail by its own terms? We interviewed a wide range of officials, including CAVC judges, VLJs, staff attorneys, and senior officials at BVA and the VA, to understand the dynamics in more detail. One possibility is that CAVC decisionmaking may itself be hard to predict.³³⁸ If so, a QR attorney simply could not predict which cases would get remanded if appealed. By this account, the failure of QR at BVA does not reflect poorly on

³³⁸ See, e.g., Telephone Interview with Interview Subject No. 1 (Aug. 27, 2018) (“It really is very judge dependent . . . at the Court of Appeals for Veterans Claims.”).

BVA but rather on CAVC. The QR experience might hence illuminate the futility of an accuracy measure pegged to expected outcomes on appeal. Yet the data refute this argument. When the QR team identified erroneous BVA decisions, those decisions were also more likely to be reversed and remanded by CAVC.³³⁹

Another explanation shifts the spotlight to VLJs. Perhaps they simply ignore exception memoranda and sign decisions with errors likely to prompt remands by CAVC. This hypothesis hardly presents BVA in a favorable light. The QR program costs BVA about \$780,000 in salaries alone each year.³⁴⁰ Moreover, QR personnel, generally thought to be among the agency's most successful staff attorneys, were taken off decision-writing teams for the detail.³⁴¹ Given this commitment of financial and attorney resources by an agency struggling with backlogs, routine VLJ indifference to exception memoranda would indict BVA for a serious management failure. But our interviews do not corroborate this hypothesis. According to one official, VLJs "almost always change" their drafts upon receiving an exception memorandum,³⁴² this impression squares with the institutional knowledge one of us acquired during his twelve years at BVA. Moreover, even if some VLJs were indeed indifferent to exception memoranda, this indifference surely would not have been uniform across all VLJs. Presumably at least some of them valued QR feedback, wanted to avoid mistakes, and gratefully incorporated suggested changes before saddling veterans with erroneous decisions.³⁴³ We would thus expect to see statistically significant differences in the effect of QR from VLJ to VLJ. But the data reveal no evidence of such heterogeneous effects across VLJs. The VLJ indifference hypothesis also does not explain why cases given a clean bill of health by the QR team are still reversed and remanded by CAVC roughly three-quarters of the time when they are appealed.

This latter fact suggests the most compelling explanation for the program's inefficacy: the standard of review. Formally, QR attorneys were supposed to call errors based upon their prediction of CAVC outcomes. Yet, likely due to pressures to allow the Chairman to annually report high accuracy

339. Ho et al., *supra* note 330, at 257.

340. *Id.* at 247.

341. *See id.*

342. Telephone Interview with Interview Subject No. 4, *supra* note 5; *see also* Telephone Interview with Interview Subject No. 2, *supra* note 110 (stating that VLJs formally contested a "single digit number" of exception memoranda each year); Telephone Interview with Interview Subject No. 3, *supra* note 33 ("[M]ost of the time they did make the change.").

343. *Cf.* Telephone Interview with Interview Subject No. 1, *supra* note 338 ("[S]ome [VLJs] welcomed . . . correcting a decision that was not legally correct before it went out.").

rates, the QR program came to grade VLJ decisions on a more forgiving curve. While the program likely corrected some errors, its relative leniency meant that a lot of cases destined for CAVC remands met the error-free standard as QR attorneys applied it. Several pieces of evidence support this explanation. First, when we examine variation in QR attorney citation rates of error, we see that higher error rates correlate with higher CAVC reversal and remand rates.³⁴⁴ The more stringent a standard of review a QR attorney used, in other words, the more her standard aligned with CAVC's.

Second, we compare the types of errors that attorneys called with the reason for CAVC remands. By far the largest number of CAVC remands involve a VLJ's failure to explain the "reason or basis" of an opinion.³⁴⁵ This is veterans' law jargon for administrative law's "deeply embedded" obligation that an agency provide a "reasoned explanation" for its action.³⁴⁶ QR attorneys called reasons-and-bases errors in under 5% of cases they reviewed.³⁴⁷ This rate increased to 10% for cases these attorneys reviewed that veterans subsequently appealed to CAVC (a difference that further corroborates that QR attorneys could meaningfully identify lower-quality cases).³⁴⁸ By contrast, CAVC remanded 62% of appeals for inadequate "reasons or bases."³⁴⁹

A 2010 unpublished memorandum written by BVA's Deputy Vice Chairman confirms what these numbers suggest, namely that QR attorneys used a far less stringent metric to judge the adequacy of a decision's "reasons or bases." The Deputy Vice Chairman wrote a memorandum to defend the FY 2009 accuracy rate the QR program generated after the VA's Office of General Counsel expressed "reservations regarding whether external data support" it.³⁵⁰ The Office had pointed out that the QR program estimated errors in 2,928 cases VLJs decided in FY 2009—1,797 fewer cases than veterans appealed to CAVC that year, and 170 fewer than the number CAVC

344. In the vast majority of instances, the QR program assigned cases to attorneys in chronological order. In other words, when an attorney was ready to receive her next decision, she simply took the decision next up in the QR queue. No QR attorney, then, was more or less likely to encounter an error. This random assignment enabled us to measure the stringency of each QR attorney's standard of review. In rare instances, the chief of the QR office would make an adjustment to this workflow to ensure balance of workloads across attorneys.

345. Ridgway & Ames, *supra* note 149, at 314.

346. Daniel A. Farber & Anne Joseph O'Connell, *The Lost World of Administrative Law*, 92 TEX. L. REV. 1137, 1152 (2014). *See generally* Ridgway & Ames, *supra* note 149 (providing an overview of reasons-or-bases remands and a critique of them).

347. Ho et al., *supra* note 330, at 259, 260 fig.5.

348. *Id.*

349. *Id.*

350. Memorandum from Steven L. Keller, *supra* note 314, at 1 (quoting the Office of General Counsel's inquiry).

remanded.³⁵¹ The Deputy Vice Chairman noted the following about “reasons or bases” remands:

While some of [the “reasons or bases”] remands are necessary, in many instances the Court remands the case for an additional explanation as to matters that are not essential to the final factual or legal conclusions already reached by the Board or matters as to which the Board’s views reasonably may be discerned from its decision. A “reasons and bases” remand does not necessarily indicate [that] the Board committed a legal or factual error, but may only reflect the Court’s judgment that an amplified explanation is desired with respect to a particular issue. The “reasons and bases” standard is both highly subjective and inconsistently applied. Therefore, it is often difficult to predict when the Court will remand a case for this reason.³⁵²

This statement is remarkable in several respects. First, it suggests BVA’s disavowal of an obligation to provide reasoned explanations for its decisions—a core duty for any agency and one central to the “legal legitimacy” of administration.³⁵³ To the extent that decisional quality includes more than accuracy, BVA’s skepticism about reason-giving is troubling. Second, the statement betrays BVA’s deviation from its own error standard. The QR program did not define errors in terms of whether the decision correctly or incorrectly awarded benefits, but in terms of expected outcomes at CAVC. How best to define an error standard may be difficult,³⁵⁴ but BVA had publicly committed to a CAVC-centered one. To then insist that the QR program should ignore the most frequent grounds for CAVC reversal ensured an inflated accuracy rate.

Third, our interviews corroborate the standard-of-review explanation and describe informal pressures to soft-pedal the QR process. As one official put it most pointedly:

The standards of quality review . . . have loosened up over time. The standard that we used to be following was essentially [whether CAVC] would be likely to vacate and remand [We] moved off of that standard quite some time ago to a lesser standard.³⁵⁵

Another told us that “the QR office knew that it was much less strict than the [CAVC], and that [the office’s] standard was much more generous than what the [CAVC] was doing.”³⁵⁶ A third expressed the difficulty of reviewing VLJ work in the role of a staff attorney, who might later report to VLJs. This

351. *See id.* at 1-2.

352. *Id.* at 2.

353. *See* Jerry L. Mashaw, Lecture, *Small Things Like Reasons Are Put in a Jar: Reason and Legitimacy in the Administrative State*, 70 *FORDHAM L. REV.* 17, 18 (2001).

354. *See infra* Part V.B.4.

355. Telephone Interview with Interview Subject No. 4, *supra* note 5.

356. Telephone Interview with Interview Subject No. 2, *supra* note 110.

person linked pressure not to call too many errors with an attorney's time-limited stint on the QR detail:

It was informal information . . . [a]long the lines of . . . “this is just a detail and you’re going to be coming back to the decision teams [supervised by VLJs.] [T]hen you’re not going to be very popular . . .” That could impact your career going forward.³⁵⁷

“There was a sense of do you want to poison the well,” this official continued: “[I]f you do your job too well, you’re going to anger or upset” VLJs.³⁵⁸

Fourth, the evolving standard of review is consistent with BVA’s decision in 2017 to jettison a QR error standard pegged to predictions of CAVC outcomes in favor of the more explicitly lenient “clear and unmistakable” standard.³⁵⁹ Just as BVA’s backlog exploded, it codified the extremely deferential standard. This change gives ample reason to doubt the integrity of the currently reported accuracy rate, and it suggests how BVA prioritizes the appearance of accuracy over more rigorous evaluation of decisional quality.

* * *

Our analysis shows that there is no reason to believe that the QR program improved the overall accuracy of VLJ decisionmaking. As one official confided to us, the open secret was that “[e]veryone . . . knows it’s kind of a sham.”³⁶⁰ Both quality assurance initiatives and individual procedural rights can yield error avoidance or error correction in particular cases. The former’s supposed institutional advantage lies with its capacity to improve decisional quality systemically. Yet the lenient standard of review QR attorneys used changed the program from one of quality improvement to one of public relations.

There is little evidence of systemic quality improvement. Given the rate at which QR attorneys called errors, a VLJ who decided 700 cases annually could expect an exception memorandum for fewer than 1% of her decisions.³⁶¹ Not only did the program generate little targeted feedback for VLJs on their specific error tendencies, it also gave them little incentive to improve. It comes as little surprise that BVA has not factored QR results into evaluations of VLJ

357. Telephone Interview with Interview Subject No. 3, *supra* note 33.

358. *Id.*

359. See Office of Quality Assurance, *supra* note 315; Office of Quality Assurance, *supra* note 325.

360. Telephone Interview with Interview Subject No. 7, *supra* note 321.

361. If about 6% of a VLJ’s decisions were sampled for QR, and if QR attorneys wrote staff attorneys regarding about 7% of the cases they reviewed, then the VLJ would receive about three exception memoranda per year. Of course, exception memoranda were presumably not evenly distributed, but even a VLJ who received twice the number of exception memoranda than the average would receive them in less than 1% of decisions.

performance.³⁶² A VLJ confronted with an exception memorandum could point to the dozens of his decisions given a clean bill of health and reject the error as nothing more than a one-off mistake.

On the other hand, there is abundant evidence that the high accuracy rate was useful as a matter of public relations. To judge by its performance before CAVC, BVA struggles to decide cases accurately. By one measure, the government's position is not "substantially justified" in as many as 70% of CAVC cases, a figure Chief Justice Roberts found "really startling" during an oral argument in 2010.³⁶³ The QR program generates a helpful counterweight, one BVA deploys to its advantage.³⁶⁴ Before Congress and in BVA's budget requests, BVA officials repeatedly tout the high accuracy rate as demonstrating how well the agency operates. In an oversight hearing by the House Veterans' Affairs Committee in 2007, for instance, the BVA Chairman described the QR system as a "fine program," boasted that GAO had "applauded our program and said it was actually exemplary," and reported a "93 percent error-free rate."³⁶⁵ In fact, even after an internal study found that the accuracy rate "will be difficult to justify as a valid measurement if confronted by external stakeholders,"³⁶⁶ the BVA Chairman reported at a House Veterans' Affairs Committee oversight hearing in 2018 that BVA's accuracy rate was "over 92%."³⁶⁷ This is what the accuracy number is for, an official remarked: It "goes in the annual report" BVA submits to Congress, and thus the number "has to be as good as possible."³⁶⁸

V. Implications

BVA's QR program is a case study in failure. It and the checkered history of quality assurance at SSA and EOIR have implications for theories of internal administrative law, particularly for scholarship that attempts to describe the

362. Telephone Interview with Interview Subject No. 3, *supra* note 33 ("I was never aware [of QR results factoring into performance evaluation.] . . . and that was something that always bugged me.").

363. See Transcript of Oral Argument at 42, 51-52, *Astrue v. Ratliff*, 560 U.S. 586 (2010) (No. 08-1322), 2010 WL 603696.

364. One official agreed that the accuracy rate "is kind of an effective way" for BVA "to respond" when it gets public criticism for poor-quality decisionmaking. Telephone Interview with Interview Subject No. 7, *supra* note 321.

365. *Hearing: BVA Adjudication*, *supra* note 314, at 30, 32 (statement of James P. Terry, Chairman, Board of Veterans' Appeals).

366. Office of Quality Assurance, *supra* note 315.

367. *Is VA Ready for Full Implementation of Appeals Reform? Hearing Before the H. Comm. on Veterans' Affairs*, 115th Cong. 15 (2018) [hereinafter *Hearing: Is VA Ready*] (testimony of Cheryl L. Mason, Chairman, Board of Veterans' Appeals).

368. Telephone Interview with Interview Subject No. 7, *supra* note 321.

relationship between internal and external sources of legal governance. This scholarship tends to criticize judicial and congressional interventions for their interference with internal agency attempts to achieve consistent, reasoned decisionmaking. It has neglected the pressures that can lead internal administrative law into dysfunction. Experiments with quality assurance help identify and categorize these pressures. They also beg the question whether a more nuanced relationship between external and internal law than what the existing literature describes might produce better overall results.

We also discuss policy implications that are a first effort to grapple with the challenges of designing institutions that meaningfully contend with quality. We suggest several interventions that can improve an agency's quality assurance program, perhaps the best chance agencies have to achieve their constitutional obligation to pass an adequate threshold of decisional quality. These interventions include not only best practices for agency governance, but also legal changes that legislatures and courts can make.

A. Theoretical Implications

To begin, we consider the implications for recent theoretical work on "internal administrative law." Such scholarship has critiqued administrative law as having too formalistic a focus on external administrative law (i.e., judicial review, congressional oversight). The irony is that despite this work's attempt to highlight the importance of internal legal sources (e.g., guidance documents), our evidence shows that it remains still too formalistic, failing to conceptualize key internal institutional dynamics that can thwart internal administrative law.

1. Internal administrative law and the threat from outside

Gillian Metzger and Kevin Stack describe the relationship between external and internal administrative law as one whereby the former tends to crowd out the latter. The more "lawlike or manifestly binding [an] agency's self-regulation is," the more likely a court is to transform that regulation into an externally enforceable obligation or to invalidate it.³⁶⁹ The paradigmatic example is the guidance document the Obama Administration crafted to afford undocumented parents of U.S. citizen children relief from deportation.³⁷⁰ The Department of Homeland Security used this document to achieve consistency in the exercise of discretion among frontline immigration officers, by prompting them to make certain decisions when confronted with certain classes of immigrants. To the extent that the guidance document bound

369. Metzger & Stack, *supra* note 182, at 1278.

370. *See id.* at 1240-41, 1240 n.2.

frontline officials, this feature made it lawlike and a form of internal governance. But the feature also sealed the document's doom. Declaring that it indeed bound officials, the Fifth Circuit held that the guidance document was really a legislative rule and thus procedurally invalid for its failure to go through notice and comment.³⁷¹ The result of such external intrusions into internal governance, Metzger and Stack argue, is "incentives for agencies to be less specific, less decisive, and less clear in their internal documents"—that is, to create less internal administrative law.³⁷² Metzger and Stack concede that, in theory, Congress or the courts could govern agencies in a way that facilitates the development of sound internal governance.³⁷³ But they are skeptical that this result will obtain in practice.³⁷⁴

Mashaw's doubt about a copacetic relationship between external and internal law, especially for mass adjudication, runs even deeper. He concludes his canonical account of disability benefits adjudication not only with optimism about SSA's capacity to govern itself justly, but also convinced that "[t]he external legal order" poses "*the threat*" to this outcome.³⁷⁵ The Due Process Revolution, for instance, forced SSA to accommodate a model of justice rooted in individual participation and desert, one epitomized by the sorts of processes that litigants enjoy in court.³⁷⁶ The result was a "tale of stress and woe,"³⁷⁷ with a shift of significant control over disability benefits adjudication to a cadre of Article III-like officials who resist managerial control and fall prey to inefficiency.³⁷⁸ Reflecting more generally on external legal control of disability benefits adjudication, Mashaw describes possible interventions as "essentially bankrupt."³⁷⁹ Congress and the courts should leave the agency to govern itself.³⁸⁰

371. *Texas v. United States*, 809 F.3d 134, 170-78 (5th Cir. 2015), *aff'd by an equally divided Court*, 136 S. Ct. 2271 (2016). For criticism of the Fifth Circuit's findings in this regard, see Blake Emerson, *The Claims of Official Reason: Administrative Guidance on Social Inclusion*, 128 YALE L.J. 2122, 2141-56 (2019).

372. Metzger & Stack, *supra* note 182, at 1288.

373. *Id.* at 1291, 1295-1296. Indeed, in another work, Metzger argues for a judicially enforceable constitutional "duty to supervise" that would entail some sort of judicially administered structural remedy devised to improve agency processes at a wholesale level. Metzger, *supra* note 137, at 1915.

374. Metzger & Stack, *supra* note 182, at 1288-89, 1292, 1296.

375. MASHAW, *supra* note 19, at 222 (emphasis added).

376. *See id.* at 29-31, 192-93.

377. *Id.* at 41.

378. *See id.* at 190-92.

379. *Id.* at 192.

380. *See Merrill*, *supra* note 14, at 53.

2. Internal pressures and distortions

While enumerating how external oversight and review can distort internal governance, leading accounts have spent less time identifying internal pressures that can distort internal administrative law.³⁸¹ Ironically, Metzger and Stack's optimism for internal administrative law may reflect an insufficiently internal inquiry into relevant governance strategies. To them, guidance documents and other formal expressions of internal governance—material readily observable outside the agency—constitute this corpus. But this sort of instrument's capacity to generate reasoned, consistent decisionmaking depends as much on agency culture, agency personnel, the agency's political environment, and other harder-to-observe variables as on the form that it takes.

The contrast between the standard of review as written and the standard of review BVA attorneys actually used for QR nicely illustrates how internal administrative law in action can diverge from internal administrative law in books. While “[s]tandards of review are not precision instruments,”³⁸² they “express[] a mood,” as Justice Felix Frankfurter insisted, and “that mood must be respected.”³⁸³ Our findings lend weight to a classic critique of these standards: that they “have no more substance at the core than a seedless grape.”³⁸⁴ BVA attorneys and CAVC judges—each experts in veterans' law—differed considerably in their decisions while ostensibly applying the same standard of review.³⁸⁵ The institutional context appears to have driven attorneys' application of the law.

Quality assurance is a useful example of internal administrative law in action because experiences with it owe their success, or failure, as much to context as to the merits of some formally stated policy. The initiatives described here illuminate at least three pressures that can distort a regime of

381. For limited discussion of obstacles that can complicate the development of internal administrative law in this scholarship, see MASHAW, *supra* note 19, at 158-163; and Metzger & Stack, *supra* note 182, at 1303-04.

382. Kristin E. Hickman & Matthew D. Krueger, *In Search of the Modern Skidmore Standard*, 107 COLUM. L. REV. 1235, 1250 (2007).

383. *Universal Camera Corp. v. NLRB*, 340 U.S. 474, 485-87 (1951).

384. See Ernest Gellhorn & Glen O. Robinson, *Perspectives on Administrative Law*, 75 COLUM. L. REV. 771, 780 (1975); see also David Zaring, *Reasonable Agencies*, 96 VA. L. REV. 135, 153-54 (2010) (arguing against “doctrinal formalism” in the administration of standards of review and suggesting that judges review agencies' decisionmaking with roughly the same standard regardless of its formal articulation); cf. *United States v. McKinney*, 919 F.2d 405, 422-23 (7th Cir. 1990) (Posner, J., concurring) (discussing the difficulties in distinguishing standards of review more generally).

385. See *supra* Part IV.

internal administrative law: resistance problems, career and agency incentives, and conflicting goals.

a. Resistance problems

Those governed by a regime of internal law may resist its control. ALJs responded to SSA's early experiments with lawsuits, obtaining settlements and eventually rules that handcuffed the agency's power to use its programs to target low-performing adjudicators. Formal accountability for low-quality decisionmaking remains extremely rare to this day. In recent years, for instance,³⁸⁶ SSA has imposed formal discipline in several instances based on ALJ "workload performance," but each has involved case processing, not decisional accuracy.³⁸⁷ VLJs and IJs have less law they can use to shield themselves from efforts to control their decisionmaking.³⁸⁸ QR at BVA, however, suggests that agency culture may provide as solid a foundation for resistance as formal legal authority.³⁸⁹ VLJs are judicial officers, and they stand higher on the VA's bureaucratic ladder than QR personnel. Both are institutional facts that may prompt resistance.³⁹⁰ One official involved with QR recalled a conversation about the program with a VLJ early in his tenure. "The thing you need to understand," this VLJ told the official, hinting at how much feedback the VLJ was willing to receive, "is that I have a piece of paper signed by the President on my wall and you do not."³⁹¹

386. On the paucity of disciplinary action in the 1970s and 1980s, see, for example, *Hearing: Delays*, *supra* note 199, at 81 (noting in 1975 that no ALJ had ever been removed "for lack of productivity" or because of "quality of adjudication"); and Jeffrey S. Lubbers, *The Federal Administrative Judiciary: Establishing an Appropriate System of Performance Evaluation for ALJs*, 7 ADMIN. L.J. AM. U. 589, 598-600 (1993). *Cf.* *Shapiro v. Soc. Sec. Admin.*, 800 F.3d 1332, 1338 (Fed. Cir. 2015) (commenting on the "virtually insurmountable" barrier to ALJ removal (quoting Lubbers, *supra*, at 600)).

387. See OFFICE OF THE INSPECTOR GEN., SOC. SEC. ADMIN., A-12-11-01138, OVERSIGHT OF ADMINISTRATIVE LAW JUDGE WORKLOAD TRENDS 15, E-1 & fig.E-1 (2012), <https://perma.cc/M393-XF9F>; *cf.* U.S. GOV'T ACCOUNTABILITY OFFICE, *supra* note 248, at 33 (describing "non-disciplinary" action SSA has taken to improve performance and noting that "[n]early all" such directives SSA has issued to ALJs involve "timeliness").

388. See *supra* Parts I.A.2-3.

389. On culture and resistance, see Nina Rabin, *Victims or Criminals? Discretion, Sorting, and Bureaucratic Culture in the U.S. Immigration System*, 23 S. CAL. REV. L. & SOC. JUST. 195, 225-26 (2014).

390. *Cf.* William H. Simon, *Where is the "Quality Movement" in Law Practice?*, 2012 WIS. L. REV. 387, 402-03 (describing resistance by professionals, including lawyers, to performance measurements).

391. Telephone Interview with Interview Subject No. 2, *supra* note 110; see also Telephone Interview with Interview Subject No. 3, *supra* note 33 (noting that VLJs, "as they would say, were approved by the President and who the heck was I to be saying that they were wrong"). As noted in Part IV, officials told us that VLJs generally revised opinions after receiving exception memoranda. This practice does not rebut concerns about
footnote continued on next page

Recognizing that officials subject to an internal governance regime may resist, the agency may design the regime accounting for this resistance, possibly at the expense of sound administration. The highly deferential standard of review SSA adopted for its first quality assurance program, for example, resulted in part from the agency's concern that a more searching one would provoke ALJ backlash. BVA's official standard of review, adopted in 2002, asked reviewers to predict how a particular case would fare at CAVC if appealed. According to one person familiar with the program, however, reviewers ceased attempting to predict CAVC outcomes within a decade.³⁹² This evolution fits our findings: that quality reviewers used a more deferential standard than did CAVC notwithstanding the stated equivalence between the two.³⁹³

One possible reason for this deference has to do with the strained relationship between BVA and CAVC. No CAVC judge has ever served at BVA and thus none has ever experienced firsthand BVA's docket and processes.³⁹⁴ As a result, one official told us, "there's definitely an attitude [among VLJs], and I think very, very much justified, that [CAVC judges] have no idea what's going on."³⁹⁵ "I personally have no respect for the [CAVC]," an official declared, explaining that its judges "nitpick everything."³⁹⁶ A standard of review that rigorously graded the quality of VLJ decisions by a metric that predicted what CAVC judges would do might have prompted resentment, and then resistance, by VLJs.

b. Career and agency incentives

Incentives can operate at both the individual official level and the agency level to distort regimes of internal administrative law. A regime can fail if officials responsible for its administration have an incentive to administer the regime in a manner that departs from its stated purpose. BVA staff attorneys, for example, may have had an incentive to review VLJ decisions lightly. As one

resistance at BVA. It is possible that the QR program ensured that VLJs would accept exception memoranda so long as the program issued few of them. By anticipating and dodging VLJ resistance, in other words, the QR program could maintain the effectiveness of the relatively small number of exception memoranda.

392. Telephone Interview with Interview Subject No. 1, *supra* note 338.

393. *See also Hearing: BVA Adjudication*, *supra* note 314, at 39 (statement of Barton F. Stichman, Joint Executive Director, National Veterans Legal Services Program) ("There obviously is a major disconnect between the annual report card prepared by the Court of Appeals for Veterans Claims and the annual report card prepared by Board management.").

394. Telephone Interview with Interview Subject No. 2, *supra* note 110.

395. *Id.*

396. Telephone Interview with Interview Subject No. 1, *supra* note 338.

official told us, “everybody . . . who had ambitions of becoming selected to be a judge one day . . . competed for” selection for a QR detail.³⁹⁷ In fact, of the forty-three staff attorneys who served on QR details from April 2002 to November 2016, nineteen have become VLJs and another five have become Senior Counsel, a highly competitive position and stepping stone to a VLJ appointment.³⁹⁸ Knowing that she is on track to win further promotion, a staff attorney assigned to QR might, whether consciously or not, worry about angering VLJs—a constituency with influence over her future³⁹⁹—by administering too searching a standard of review.⁴⁰⁰ Even if a staff attorney does not harbor judicial ambitions, she might still review decisions lightly. The time-limited nature of a QR detail means that the attorney will cycle back to writing decisions under the supervision of a VLJ, an official who can make or break her job.⁴⁰¹

If an incentive to defer does affect BVA QR, it results from a long-recognized and obvious design flaw—the use of bureaucratic subordinates to evaluate the work of bureaucratic superiors.⁴⁰² But the flaw has persisted for two decades,⁴⁰³ a stubbornness that may indicate the operation of a different

397. Telephone Interview with Interview Subject No. 2, *supra* note 110.

398. One of us with substantial experience at BVA compiled these figures, using his own institutional knowledge supplemented by information from BVA annual reports and internal agency documents obtained through FOIA.

399. For a description of the VLJ appointments process, see *The Evolution of the Veterans’ Appeals Process: A Conversation with Cheryl Mason* at 52:02-54:23, FED. NEWS NETWORK: BUS. GOV’T HOUR (Mar. 14, 2019, 11:30 AM), <https://perma.cc/XC3K-A4CD>. VLJs can affect attorneys’ prospects in several ways. They write performance evaluations for attorneys, for instance, and may be consulted when considering staff attorneys for VLJ openings. VLJs can also control the workflow for attorneys by assigning more difficult or easier cases, affecting the attorney’s ability to hit annual production targets. Telephone Interview with Interview Subject No. 4, *supra* note 5.

400. Telephone Interview with Interview Subject No. 3, *supra* note 33 (“You want to do your job, but if you do your job too well, you’re going to anger or upset [VLJs]. That could have a significant impact on your career.”). *But see* Telephone Interview with Interview Subject No. 5 (Sept. 17, 2018) (disagreeing with the premise that concern for career prospects impacted the QR process).

401. One official described to us how a VLJ needs to manage his docket with care to ensure that all staff attorneys under the VLJ’s supervision can meet their annual production quotas. Telephone Interview with Interview Subject No. 4, *supra* note 5.

402. *Cf.* MASHAW, *supra* note 190, at 176 (insisting that “considerable care must be taken to ensure the independence of the quality assurance staff”).

403. In its 2002 audit, the General Accounting Office noted that BVA’s program violated “the governmental internal control standard calling for separation of key duties and the governmental performance audit standard calling for organizational independence for agency employees who review and evaluate program performance.” U.S. GEN. ACCOUNTING OFFICE, *supra* note 305, at 7 n.7. While BVA responded to this critique by ensuring that QR members were separated in the organizational hierarchy from the decision-writing teams, *see id.*, the employment dynamics we highlight still hold. Most
footnote continued on next page

set of incentives. Depending on how the errors skew, a malfunctioning QR program may actually serve the agency's political interests. An entrenched political consensus favors generosity in the veterans' benefits system.⁴⁰⁴ BVA may thus face no political consequences if its adjudicators tend to err in veterans' favor. Immigration adjudication suggests a similar dynamic. EOIR's failure to adopt any meaningful quality assurance initiatives for immigration courts in the face of longstanding criticism of their systemic accuracy makes no sense if the agency prioritizes sound IJ administration of immigration policy. But it does jive with political forces if IJs tend to err in ways that disfavor immigrants. An increased pace of deportations fits the pro-enforcement preferences of some recent administrations and the institutional commitments of the Attorney General as the country's chief law enforcement officer.⁴⁰⁵

Politics in the form of interbranch relations may disincentivize improvement as much as politics in the form of ideological commitment. The Government Performance and Results Act (GPRA) requires federal agencies to list performance goals in annual reports presented to Congress, and to identify and discuss performance measures to indicate progress toward those goals.⁴⁰⁶ Ostensibly an accountability mechanism, the GPRA, as Sidney Shapiro and Rena Steinzor argue, gives agencies an incentive to devise performance goals and metrics to "ensure that they can 'pass' their own grading criteria."⁴⁰⁷ BVA's accuracy rate, never dipping below 89% from 2003 to 2016 regardless of its

importantly, staff attorneys who are subordinate to VLJs, and who cycle back to decision-writing teams after serving on a QR detail, still review VLJ work. Office of Quality Assurance, *supra* note 319, at 6 (noting that "VLJs only lose top performing attorneys for a short period" and "[a]ttorneys return to their former VLJs after detail is completed").

404. Cf. Linda D. Jellum, *Heads I Win, Tails You Lose: Reconciling Brown v. Gardner's Presumption That Interpretive Doubt Be Resolved in Veterans' Favor with Chevron*, 61 AM. U. L. REV. 59, 64 (2011) ("Congress specifically included a number of statutory advantages to veterans to ensure the nonadversarial and pro-claimant character of the administrative process.").

405. See 2 COMM'N ON IMMIGRATION, *supra* note 8, at 1-25 to -26 (discussing the Trump Administration's pro-enforcement policy goals); COMM'N ON IMMIGRATION, AM. BAR ASS'N, REFORMING THE IMMIGRATION SYSTEM 2-24 to -25 (2010), <https://perma.cc/H4RH-WF6B>; Stephen H. Legomsky, *Deportation and the War on Independence*, 91 CORNELL L. REV. 369, 372-75 (2006) (discussing the George W. Bush administration's pro-enforcement goals).

406. CLINTON T. BRASS, CONG. RESEARCH SERV., R42379, CHANGES TO THE GOVERNMENT PERFORMANCE AND RESULTS ACT (GPRA): OVERVIEW OF THE NEW FRAMEWORK OF PRODUCTS AND PROCESSES 5 (2012), <https://perma.cc/ADH9-A9SL>; see also Government Performance and Results Act of 1993, Pub. L. No. 103-62, 107 Stat. 285 (codified as amended in scattered sections of the U.S. Code).

407. Sidney A. Shapiro & Rena Steinzor, *Capture, Accountability, and Regulatory Metrics*, 86 TEX. L. REV. 1741, 1744 (2008).

caseload,⁴⁰⁸ certainly seems like “a sunny set of invented statistics designed to reassure [its] overseer[] that [it is] doing fine.”⁴⁰⁹ Congress has little ability to dissect the process generating these numbers or to conduct its own quality audit of something as complex as veterans’ benefits decisionmaking. It effectively has to take the results as true.⁴¹⁰ BVA has every incentive to continue with a program that helps it with congressional relations, however ineffective it is as a strategy to protect decisional quality.

c. Conflicting goals

Agencies handling anything more than a modestly complicated task will have multiple, potentially conflicting goals to pursue.⁴¹¹ When agencies are left to govern themselves, at least three related phenomena may result. First, the agency may prioritize which goals it pursues in a manner that best serves its institutional interests and not necessarily the interests of its constituents or the larger public good.⁴¹² Second, the agency may prioritize goals that have straightforward performance metrics over others whose measurement is more difficult.⁴¹³ Third, the agency may devise metrics or ways of presenting data about its accomplishments with its own short-term interests in mind, not necessarily to present the most accurate portrait of its achievement.⁴¹⁴

The enduring conflict between quality and quantity illustrates these phenomena and how they can distort a regime of internal law. Because an agency can more readily measure the number of cases it decides than decisional quality,⁴¹⁵ it will tend to favor the pursuit of quantitative goals over qualitative

408. *See supra* note 313.

409. Shapiro & Steinzor, *supra* note 407, at 1744 (describing how the GPRA encourages agencies to use favorable statistics to protect themselves).

410. Telephone Interview with Interview Subject No. 4, *supra* note 5 (describing the complexities of veterans’ law and the difficulties they pose for congressional oversight).

411. Donald P. Moynihan, *The Promises and Paradoxes of Performance-Based Bureaucracy*, in THE OXFORD HANDBOOK OF AMERICAN BUREAUCRACY 278, 286 (Robert F. Durant ed., 2010); *see also* Eric Biber, *Too Many Things to Do: How to Deal with the Dysfunctions of Multiple-Goal Agencies*, 33 HARV. ENVTL. L. REV. 1, 10 (2009) (noting in an analysis of government agencies and goals that the notion that a principal only has one goal it assigns to an agent to achieve “is rarely reflective of reality”).

412. *See* Moynihan, *supra* note 411, at 286-87.

413. *Id.* at 288; *see also* Biber, *supra* note 411, at 12.

414. Moynihan, *supra* note 411, at 286-87.

415. *Cf.* MASHAW, *supra* note 19, at 158-60 (discussing the “[n]ormative [a]mbiguity” of defining successful performance).

ones.⁴¹⁶ An agency may therefore sacrifice its quality assurance initiatives when backlogs increase. SSA's first experiment with "own motion" review ended in 1975 when the agency, facing an exploding docket, reassigned staff administering the program to case production.⁴¹⁷ More recently, as backlogs surged and ALJ productivity declined,⁴¹⁸ SSA suspended its inline quality reviews and disability case reviews and reallocated staff assigned to them to deciding cases.⁴¹⁹ BVA did not respond to the spike in its caseload by increasing its QR program's capacity. Instead, it effectively halved the percentage of cases the program sampled for review and reduced the number of attorneys assigned to the office.⁴²⁰

BVA's behavior over the past decade exemplifies how an agency's institutional interests affect its goal prioritization. A BVA official conveyed that Congress's "concern is more focused on 'what are you doing to get rid of the backlog'" than quality.⁴²¹ Even so, presumably a dramatic drop in the self-reported accuracy rate would spark oversight concerns.⁴²² If rising caseloads do impact quality, then this sort of drop should happen—unless BVA grades itself on a more forgiving curve. Just this sort of inflation appears to have happened. The shift to the "clear and unmistakable error" standard and the shortening of QR positions from two years to six months, whatever their legitimate motivations, have helped BVA continue to report a very high accuracy rate even as it maintains what one congressperson called an "absolutely alarming" pace of decisionmaking.⁴²³

416. Moynihan, *supra* note 411, at 288 ("As a result [of an agency's emphasis on readily measured goals], efficiency goals often are pursued at the expense of program quality . . .").

417. *See Hearing: ALJ Role*, *supra* note 227, at 65 (statement of Charles N. Bono, President, Association of Administrative Law Judges, Inc.); STAFF OF H. SUBCOMM. ON SOC. SEC., *supra* note 115, at 25.

418. *E.g.*, OFFICE OF THE INSPECTOR GEN., *supra* note 266, at 3-4, 4 fig.2 (describing a decrease in ALJ productivity from FY 2012 to FY 2014).

419. U.S. GOV'T ACCOUNTABILITY OFFICE, *supra* note 248, at 40.

420. *See* Office of Quality Assurance, *supra* note 319, at 2 (indicating that 136 cases would be sampled out of 6,000 in a month). This rate of 2.3% would be less than half of the 5% sampling rate for original cases in the earlier program. Based on our understanding, such a decreased workload was accompanied by a reduction in the team from seven individuals to four, including the position of Chief.

421. Telephone Interview with Interview Subject No. 5, *supra* note 400; *see also* Telephone Interview with Interview Subject No. 4, *supra* note 5 (insisting that Congress is almost exclusively interested in the number of cases decided and timeliness indicators).

422. *Cf.* Telephone Interview with Interview Subject No. 4, *supra* note 5 (commenting on congressional interest in quality and insisting that "[i]f there's that dramatic of a drop then they'd start to say something").

423. *Assessing Whether VA Is on Track to Successfully Implement Appeals Reform: Hearing Before the H. Comm. on Veterans' Affairs*, 115th Cong. 20 (2018) (Rep. Brian Mast); *see also* footnote continued on next page

As BVA's backlog has grown, the VA has subtly deemphasized quality as a key performance measure. In 1998, BVA's chairman declared that "[q]uality in appellate decision-making . . . is the Board's single most important goal."⁴²⁴ VA's annual fiscal year budget requests to Congress from 2008 to 2010 reflected this emphasis; they included a "[d]eficiency-free decision rate" as BVA's first performance measure.⁴²⁵ From 2011 to 2014, the requests stopped reporting the accuracy rate, although they did refer to "expectations for quality, timeliness, and responsiveness" as an "[o]bjective."⁴²⁶ Its requests from 2015 to 2019, when our findings became public, have omitted quality as a performance measure altogether. Instead, they refer exclusively to case production.⁴²⁷

To be sure, a preference for quantity over quality is not itself problematic. An agency transparent about the tradeoff could put the onus on Congress to respond to increasing caseloads, forcing it either to decide which goal it wants the agency to prioritize, or to commit sufficient resources so that the goals no longer conflict. But officials anxious to avoid congressional scrutiny or public outcry may prefer to soft-pedal any problem,⁴²⁸ even if a more frank acknowledgment might buttress a request for additional resources. The agency may gain power and autonomy if it can strengthen or defend its reputation for

Hearing: Is VA Ready, *supra* note 367, at 15 (testimony of Cheryl L. Mason, Chairman, Board of Veterans' Appeals) (referring to an accuracy rate of "over 92%" to tout the quality of BVA decisionmaking). Compare U.S. DEP'T OF VETERANS AFFAIRS, *supra* note 94, at 5 (projecting an "unprecedented" 81,033 decisions for FY 2018), with *id.* at 17 (reporting the FY 2018 year-to-date accuracy rate as 95%).

424. Memorandum from Richard B. Standefer, *supra* note 301, at 1.

425. 3 U.S. DEP'T OF VETERANS AFFAIRS, FY 2010 BUDGET SUBMISSION, at 5C-5 tbl.1 (2009), <https://perma.cc/LZW8-3YHF>; 3 U.S. DEP'T OF VETERANS AFFAIRS, FY 2009 BUDGET SUBMISSION, at 5C-2 tbl. (2008), <https://perma.cc/S5RJ-BC28>; 2 U.S. DEP'T OF VETERANS AFFAIRS, FY 2008 BUDGET SUBMISSION, *supra* note 313, at 7C-2 tbl.

426. 3 U.S. DEP'T OF VETERANS AFFAIRS, FY 2014 BUDGET SUBMISSION, at 4C-4 tbl.1 (2013), <https://perma.cc/YZM8-GKGE>; 3 U.S. DEP'T OF VETERANS AFFAIRS, FY 2013 BUDGET SUBMISSION, at 5C-4 tbl.1 (2012), <https://perma.cc/65W4-UMQV>; 3 U.S. DEP'T OF VETERANS AFFAIRS, FY 2012 BUDGET SUBMISSION, at 5C-4 tbl.1 (2011), <https://perma.cc/T3NG-B9HU>; 3 U.S. DEP'T OF VETERANS AFFAIRS, FY 2011 BUDGET SUBMISSION, at 5C-5 tbl.1 (2010), <https://perma.cc/KJ8N-VQUM>.

427. 3 U.S. DEP'T OF VETERANS AFFAIRS, FY 2019 BUDGET SUBMISSION, at BVA-266 to -268 (2018), <https://perma.cc/W7GY-TBPZ>; 3 U.S. DEP'T OF VETERANS AFFAIRS, FY 2018 BUDGET SUBMISSION, at BVA-296 to -299 (2017), <https://perma.cc/6MG6-8F6B>; 3 U.S. DEP'T OF VETERANS AFFAIRS, FY 2017 BUDGET SUBMISSION, at BVA-274 to -279 (2016), <https://perma.cc/HHW7-SUBC>; 3 U.S. DEP'T OF VETERANS AFFAIRS, FY 2016 BUDGET SUBMISSION, at BVA-273 to -276 (2015), <https://perma.cc/MRD2-4MUD>; 3 U.S. DEP'T OF VETERANS AFFAIRS, FY 2015 BUDGET SUBMISSION, at BVA-9 tbl.1 (2014), <https://perma.cc/3K8P-N6H7>.

428. *Cf.* Moynihan, *supra* note 411, at 288 (noting that conflicting goals and goal ambiguity will lead agencies to privilege "short-term goals over long-term measures of effectiveness").

competent performance.⁴²⁹ Put conversely, “politicians do not respond to a structure fraught with three-alarm fires by supplying more fuel; they are just as likely to . . . begin choosing other venues in which to vest authority and power.”⁴³⁰ Recent testimony by the BVA Chairman illustrates this tendency. She disclaimed a need for additional resources,⁴³¹ even as VLJ caseloads are anticipated to more than double over the next five years.⁴³²

This link between reputation and power may explain an arguably counterintuitive pattern in the history of SSA’s quality assurance programs. SSA pursued a three-pronged quality assurance initiative at times of rising backlogs, pairing a push for an increased rate of decisionmaking with centrally administered efforts to ensure that ALJs met an adequate threshold of decisional quality.⁴³³ Perhaps this coupling indicates that SSA appreciates the threat that an emphasis on quantity poses to quality, and it has attempted to mitigate the impact of faster decisionmaking on decisional accuracy. A more cynical take views SSA’s quality assurance efforts as window dressing to give the agency institutional cover as it really pursues the more easily measurable—and therefore favored—quantitative goal.

B. Improving Quality Assurance: Externally Influenced Internal Administrative Law

Dysfunction in internal administrative law is not inescapable. An agency could implement meaningful quality assurance programs without action from Congress or the courts. With little apparent prodding from outside the agency,⁴³⁴ SSA has recently experimented with an extensive set of changes to its quality assurance programs. Its chief designer claims that they have improved decisional quality even as adjudicators decide more cases.⁴³⁵ External observers still need to confirm these changes’ efficacy. But they suggest something promising.

429. See DANIEL CARPENTER, REPUTATION AND POWER: ORGANIZATIONAL IMAGE AND PHARMACEUTICAL REGULATION AT THE FDA 46 (2010).

430. *Id.* at 54.

431. *Hearing: Is VA Ready*, *supra* note 367, at 13-14 (testimony of Cheryl L. Mason, Chairman, Board of Veterans’ Appeals).

432. See *supra* text accompanying notes 94-95.

433. See *supra* Part III.A.1.

434. In fact, according to one observer, the person responsible for these changes implemented them without seeking any go-ahead from top agency officials. Telephone Interview with Interview Subject No. 2, *supra* note 110.

435. See Ray & Lubbers, *supra* note 114, at 1575 n.*, 1606; *supra* text accompanying notes 84-87.

One of us likewise spearheaded a change to BVA's QR program without any external mandate to do so.⁴³⁶ Between November 2016 and September 2017, the revised program did not generate a crude accuracy rate. Rather, BVA staff attorneys used their review of draft decisions, along with other sources of information like CAVC remands, to identify recurring problems that tend to generate a disproportionate number of errors in VLJ decisions. This information was intended to support systemic interventions, such as focused training for all VLJs or policy clarification. The first "Targeted Quality Analysis" determined that a particular requirement involving "extraschedular ratings" produced significant disparities between BVA and CAVC, even though remands almost never resulted in increased benefits for veterans. The report thus recommended that the VA replace the regulation, which simply generated more "churn," with more specific, easily applied eligibility criteria. A category of mistakes would disappear, and deserving veterans would get their benefits more efficiently. To be sure, this program was not rigorously evaluated, but if SSA history teaches us one thing, it is that such experimentation is required to crack the quantity-quality nut.

The fate of this revised program underscores the fragility of internal administrative law. When one erroneous decision caused the VA embarrassment, the BVA Chairman scuttled the program and insisted that QR go back to generating the old accuracy number.⁴³⁷ The old program only sampled about 6% of VLJ decisions for review,⁴³⁸ so in all likelihood BVA would have issued the embarrassing decision even had the newer program remained in place. But bringing back the old accuracy number might provide cover in light of criticism. This development echoes SSA's decision to reinvigorate its quality assurance efforts after reports about biased decisionmaking surfaced in 1992.⁴³⁹ The agency's action suggested its seriousness about quality, but it didn't grapple with the underlying problem that prompted criticism in the first place.

When pressures like resistance, incentives, and conflicting goals interfere with internal reform, external intervention by Congress and the federal courts may help. Our proposal below would not require an agency to do anything specific, but would leave it considerable room for internally controlled design and experimentation. We argue that external legal interventions can facilitate, rather than hinder, the development of effective internal administrative law.

436. The details in this paragraph, like those discussed in Part III.C, are based on personal knowledge and experience. *See also supra* note 291.

437. *See supra* notes 319-21 and accompanying text.

438. *See Ho et al., supra* note 330, at 251 tbl.1 (reporting that roughly 6% of cases are selected for QR).

439. *See supra* note 243 and accompanying text.

1. Mandating disclosure of methodology

Our BVA findings show how an agency can deploy an “accuracy statistic” to blunt oversight and public scrutiny. BVA has invoked that figure repeatedly to assuage overseers’ expectations that its adjudication meets an adequate threshold of systemic quality.⁴⁴⁰ To preserve the possibility of meaningful oversight, Congress should require agencies to disclose information about the methodology of QR. The fact that BVA formally advertised one standard of review (prediction of CAVC disposition) and functionally deployed a more lenient one illustrates the risk of a lack of transparency. Just as scientists cannot publish results without disclosing methodology, agencies should not be able to publish performance measures without disclosing measurement methods. Such disclosure would empower private entities—academic researchers, investigative journalists, nonprofits, and public interest organizations—to hold agencies accountable to their published accuracy metrics.

Agencies should make sufficient data public so that observers can independently assess program efficacy. Although we accessed data sufficient to evaluate BVA’s QR program, our ability to do so depended on considerable institutional insight from one of us having worked for the agency. FOIA requests involve considerable guesswork and delays, which might defeat all but the most dogged efforts. The FOIA Improvement Act of 2016 has implemented a “Rule of 3,” whereby three requests for the same disclosure trigger an agency’s duty to make the disclosure without additional requests.⁴⁴¹ Given the obvious interest in information like the VACOLS data, an agency should not wait for multiple FOIA requests to discharge this obligation.⁴⁴²

In addition, agencies should be required to disclose details about how they design and administer their quality assurance programs. The institutional setup of QR at BVA, for instance, should prompt skepticism about the accuracy rate. The use of bureaucratic subordinates to review their superiors’ work, the weakening of the standard of review, and the shortening of attorney terms on QR details all suggest that BVA grades itself on an increasingly lenient curve.

440. See, e.g., *Hearing: Is VA Ready*, *supra* note 367, at 15 (testimony of Cheryl L. Mason, Chairman, Board of Veterans’ Appeals) (referring to an accuracy rate of “over 92%”); *Hearing: BVA Adjudication*, *supra* note 314, at 30 (statement of James P. Terry, Chairman, Board of Veterans’ Appeals) (using the accuracy rate from the QR program to defend against claims that the quality of BVA’s work was suffering).

441. *OIP Summary of the FOIA Improvement Act of 2016*, U.S. DEP’T JUST., <https://perma.cc/5V5E-HTQR> (last updated Aug. 17, 2016).

442. EOIR, for instance, made data on IJ decisionmaking available only after receiving multiple FOIA requests. See *id.* (providing data “[a]s EOIR has received at least three FOIA requests for this information”).

Comprehensive disclosure would help to reduce perverse incentives to design rubber-stamp QR programs. Disclosure may also reduce the propensity by Congress to accept agencies' own measures at face value. External scrutiny may force agencies to clarify when goals conflict, and thus better force Congress to assume responsibility for articulating which goal it prioritizes and why. BVA's accuracy measure excuses Congress from the difficult task of ranking case production relative to decisional quality. Legislators can pressure the VA to reduce backlogs without any consideration for the effect increased production has on decisional quality. Agency personnel can object to the ever-increasing demands on their time. But if quality does not suffer, then the result is simply that government employees work harder, hardly an outcome any member of Congress would likely lament.⁴⁴³ If confronted with the costs that the pressure to produce can have, Congress at least may be forced to assume responsibility for goal prioritization.

2. Institutionalizing oversight

Congress should institutionalize oversight of agency QR by requiring an independent body to audit quality assurance initiatives with some regularity. The reason is that disclosure alone may provide insufficient incentives to study QR in sufficient detail.⁴⁴⁴ Disclosure alone may place too high an informational burden on Congress and the public to interpret complicated internal procedures and analyze massive amounts of data. As a result, stakeholders may simply accept the agency's characterization. We see three potential institutional avenues. First, each agency has its own Inspector General (IG). SSA's IG has evaluated some of the agency's quality assurance initiatives,⁴⁴⁵ but the VA's IG has tended to focus on picayune matters of little significance to systemic integrity.⁴⁴⁶ Leadership

443. Telephone Interview with Interview Subject No. 4, *supra* note 5 (“Congress doesn’t care if a bunch of federal bureaucrats . . . have to work 100 hours a week.”).

444. Cf. Daniel E. Ho, *Fudging the Nudge: Information Disclosure and Restaurant Grading*, 122 YALE L.J. 574, 650-56 (2012) (discussing the limitations of information disclosure in the context of consumers).

445. *E.g.*, OFFICE OF THE INSPECTOR GEN., *supra* note 266, at B-2 to -3; OFFICE OF THE INSPECTOR GEN., *supra* note 259, at 3-10.

446. *See, e.g.*, OFFICE OF INSPECTOR GEN., U.S. DEP’T OF VETERANS AFFAIRS, 15-02747-314, BOARD OF VETERANS APPEALS: REPORT OF ADMINISTRATIVE INVESTIGATION; ALLEGED PROHIBITED PERSONNEL PRACTICE (2016), <https://perma.cc/VS79-NVSD> (finding in an investigation that members of a VLJ appointment panel disclosed confidential information about their process); Memorandum from Assistant Inspector Gen. for Investigations, to Acting Chairman, Bd. of Veterans’ Appeals (May 1, 2012), <https://perma.cc/R3ER-SYDK>.

vacuums⁴⁴⁷ and funding shortfalls⁴⁴⁸ may explain how little attention the VA's IG has paid to a relatively small division within the VA and the limited scope of its investigations. Without an instruction to pay attention to decisional quality, IGs may neglect the issue.

The GAO provides another avenue. After all, the GAO has studied SSA quality assurance programs and prompted meaningful change with its critique of BVA's program in 2002.⁴⁴⁹ But the GAO takes its marching orders from congressional requesters,⁴⁵⁰ and its attention to decisional quality has, to date, remained only episodic. The GAO may nonetheless be the institution most likely to be able to conduct a rigorous accounting of the effectiveness of quality assurance initiatives.

Third, Congress could create independent oversight boards—comprised of researchers, former agency officials, representatives of program beneficiary groups, and other stakeholders—to provide periodic assessments of the status of QR. Such boards could function much like panels for the National Academy of Sciences or reports for the Administrative Conference of the United States, providing expert input and recommendations for improving quality assurance. We recognize that such oversight requires some resources. But such institutions may be necessary to counteract the pathologies we have documented.

3. Improving design of quality assurance

Congress could mandate basic design principles to align incentives and reduce resistance against quality management.

BVA's program is institutionally flawed because it tasks bureaucratic subordinates with the responsibility for reviewing the work of their bureaucratic superiors. SSA's programs, in contrast, mostly have relied on work done by Appeals Council personnel, who presumably have no professional incentive to go easy on ALJs. On the surface, the structure of QR at SSA seems superior and suggests an obvious change BVA could make. The issue of institutional design, however, raises more complex questions. ALJs can

447. See Anne Joseph O'Connell, *Vacant Offices: Delays in Staffing Top Agency Positions*, 82 S. CAL. L. REV. 913, 957 & tbl.2 (2009) (noting that agency IGs have been among the last positions to be appointed in recent administrations).

448. See *Strengthening the Unique Role of the Nation's Inspectors General: Hearing Before the S. Comm. on Homeland Sec. & Governmental Affairs*, 110th Cong. 46 (2007) (statement of Glenn A. Fine, Inspector General, U.S. Department of Justice) (claiming that IG offices' biggest challenge is underfunding).

449. See *supra* notes 306-12 and accompanying text.

450. See U.S. GOV'T ACCOUNTABILITY OFFICE, GAO-17-767G, GAO'S CONGRESSIONAL PROTOCOLS 7-8 (2017), <https://perma.cc/LJ46-44AM> (describing the GAO's work priorities).

resent direction from SSA's Appeals Council, a body that some ALJs view as arbitrary and disconnected from the day-to-day realities of benefits adjudication.⁴⁵¹ Review done by bureaucratic superiors, while blunting the incentive effect, might increase resistance. It also might deny a QR program the benefits of true peer review, in which adjudicators with the same basic job duties review each other's work.⁴⁵²

Congress does not need to specify a particular institutional structure for QR, but it could draw on existing standards to ensure that review (a) is independently conducted, and (b) provides incentives for improvement. First, results could either be used in performance evaluation of staff attorneys and/or provide adjudicators credit if they show improvement on identified areas of concern. Second, serious consideration should be given to the tradeoffs involved in determining who conducts QR. At BVA, for instance, it may be desirable to assign VLJs, as opposed to staff attorneys, the responsibility of QR. Such VLJs could be relieved of regular caseload to work as part of the QR team. Alternatively, rather than using a specialized QR unit, the process could be designed as a collaborative form of peer review to promote a culture of trust in feedback and guidance.⁴⁵³ Third, review could be anonymized to reduce interpersonal pressure. If a QR attorney at BVA, for instance, flags an error, the exception memorandum could issue in the program's name and not the attorney's. The reviewed VLJ would have no reason to resent a particular colleague.

Last, design may include reporting requirements. To the extent that the GPRA requires reporting of something like BVA's accuracy rate, it demands too crude a measure of decisional quality that is hence subject to manipulation. A reporting requirement ought to place "less emphasis on the objective of precisely measuring government performance," and instead require agencies to "develop . . . more effective policy tools for guiding program management."⁴⁵⁴ Instead, agencies could be required to collect and provide data on issues that arise systemically in decisionmaking, report steps they are taking to respond, and conduct rigorous assessments of the efficacy of interventions.

451. See GELBACH & MARCUS, *supra* note 49, at 126.

452. For more on the potential costs and benefits of peer review, see Daniel E. Ho, *Does Peer Review Work? An Experiment of Experimentalism*, 69 STAN. L. REV. 1, 79-82 (2017).

453. *Id.* at 50-52.

454. Carolyn J. Heinrich, *Outcomes-Based Performance Management in the Public Sector: Implications for Government Accountability and Effectiveness*, 62 PUB. ADMIN. REV. 712, 722 (2002) (emphasis omitted).

4. Structural reform litigation

Our discussion of Congress's role rests on an asymmetric assumption: An agency on its own lacks sufficient motivation to adopt a set of good government reforms, but the promise of good government could spur Congress to act. The politics of decisional quality can fail, of course, and a systemic skew in error patterns may cause them to do so. The total dearth of meaningful quality assurance at EOIR suggests that if errors skew at all, they probably disfavor immigrants,⁴⁵⁵ a politically marginalized population that may not easily find a sympathetic ear in Congress. Canonical accounts of judicial review justify court involvement precisely under such circumstances.⁴⁵⁶ When quality has broken down, structural reform litigation may spur an agency to reform.

Judicial intervention to protect decisional quality nonetheless prompts at least four concerns, which we address here.

(i) *Displacement of Internal Governance*. Skeptics argue that judicial review tends to harm the development of internal administrative law.⁴⁵⁷ Yet public law litigation can preserve room for internal agency governance. As William Simon and Charles Sabel argue, the prevailing approach to public law remedies has “move[d] away from command-and-control toward experimentalist methods.”⁴⁵⁸ Upon a determination of liability, the judge does not dictate remedial terms but rather “facilitat[es] a process of deliberation and negotiation among the stakeholders,” including the defendant agency itself.⁴⁵⁹ The result, a “rolling-rule regime” that “incorporate[s] a process of reassessment and revision,”⁴⁶⁰ is not intended “to coerce obedience” but rather to “induce internal deliberation and external transparency.”⁴⁶¹ A court could insist that an agency protect decisional quality, but it can let the agency, along with stakeholders,

455. Close observers have documented sharply declining rates at which IJs grant relief in favor of immigrants, at a time when the current administration consistently expresses skepticism toward immigrants. *E.g.*, INNOVATION LAW LAB & S. POVERTY LAW CTR., THE ATTORNEY GENERAL'S JUDGES: HOW THE U.S. IMMIGRATION COURTS BECAME A DEPORTATION TOOL 12 (2019), <https://perma.cc/AJT2-ZVNT>; *Findings of Credible Fear Plummet amid Widely Disparate Outcomes by Location and Judge*, TRAC IMMIGR. (July 30, 2018), <https://perma.cc/SE6Y-YCFB>. For a review of the lack of quality assurance at EOIR, see also Part III.B above.

456. *E.g.*, JOHN HART ELY, DEMOCRACY AND DISTRUST: A THEORY OF JUDICIAL REVIEW 151-54 (1980).

457. *See supra* Part V.A.1.

458. Charles F. Sabel & William H. Simon, *Destabilization Rights: How Public Law Litigation Succeeds*, 117 HARV. L. REV. 1015, 1052 (2004).

459. *Id.* at 1055, 1062, 1067-68.

460. *Id.* at 1069 (emphasis omitted).

461. *Id.* at 1071.

devise institutionally acceptable methods to do so, much like a congressionally mandated oversight board.

(ii) *Jurisdiction Channeling*. A second concern involves the related problems of jurisdiction channeling and cause of action. As described in Part II, a claim that an agency has taken inadequate steps to protect decisional quality ultimately implicates the merits of individual cases. As such, the claim may not be sufficiently “collateral” to evade jurisdiction-channeling provisions that would steer it to a forum disinclined to grant a systemic remedy.

One response is to recast the cause of action not as a claim effectively alleging a lot of individual errors but rather as something more obviously collateral to the merits of particular cases. The Information Quality Act, for example, obliges federal agencies to issue and abide by guidelines “for ensuring and maximizing the quality, objectivity, utility, and integrity of information” the agency disseminates publicly.⁴⁶² Perhaps a plaintiff could challenge the integrity of something like BVA’s accuracy number and thereby force it to improve its QR program. The question whether the Act provides a private right of action remains open, and thus far litigation to force agencies to improve their information disclosures has not fared well.⁴⁶³ But no one has yet tested a theory of the sort we propose.

Another response builds on an important development in CAVC, the jurisdictionally preferred forum for cases implicating the merits of VLJ decisions. In 2017, the Federal Circuit ruled that CAVC has authority to adjudicate class actions challenging systemic shortcomings at BVA.⁴⁶⁴ The case at issue involved allegations of unconstitutional and arbitrary delays in VLJ decisionmaking,⁴⁶⁵ a theory that may present fewer obstacles to systemic reform than one rooted in quality concerns. But the decision at least recognizes the possibility that a reviewing court is not institutionally disabled from adjudicating a case involving much more than an effort to correct an error in an individual case.

Yet another response is for the federal courts to appreciate something they have recognized in other structural reform contexts involving constitutional rights: A systemic challenge to widespread agency dysfunction is not simply a bundle of individual claims alleging specific harms. When a prisoner brings a class action to challenge a state’s failure to provide adequate medical care at its facilities, she does not purport to litigate many discrete harms to inmates. Rather, she seeks to remedy the policies and practices that create a systemic

462. 44 U.S.C. § 3516 note (2018).

463. *See, e.g., Zero Zone, Inc. v. U.S. Dep’t of Energy*, 832 F.3d 654, 678 n.25 (7th Cir. 2016).

464. *Monk v. Shulkin*, 855 F.3d 1312, 1318 (Fed. Cir. 2017).

465. *See id.* at 1315, 1317-19.

risk of harm that all inmates bear.⁴⁶⁶ Likewise, a child in foster care challenging systemic mismanagement in a class action does not litigate the state's failure to provide her with a competent caseworker. Rather, her individual experience is merely a representative manifestation of the statewide risk of harm that mismanagement creates for all children in foster care.⁴⁶⁷

The equivalent doctrinal move would cast the injury to class members not as a bundle of errors committed in individual cases, but as the undifferentiated risk of harm that each veteran, immigrant, or disabled worker faces while pursuing a claim due to inadequate quality assurance efforts at the agency. This recasting fits the *Mathews v. Eldridge* framework well. Courts would inquire into the risk of an erroneous deprivation not for a discrete claimant, but for a category of claimants entitled to a particular set of procedures. A class of claimants would not seek the correction of individual errors, but rather the more plausibly collateral remedy of better protections to lower the overall risk of harm.

(iii) *Informational Barriers*. Another critique is that the informational barriers for a court to assess systemic error are too severe. To judge an agency's quality assurance efforts, one presumably needs to know the propensity of a system to err, and thus what counts as an error in the first place. A court could plausibly find an agency liable for a due process violation without defining an error standard if that agency—EOIR, for instance—makes no use of quality assurance at all. Courts typically adjudicate procedural due process challenges and the alleged benefits of additional process using logic and impressionistic understandings of a system's complexities and limitations. They tend not to require rigorous empirical proof before doing so. But a challenge to an existing quality assurance program would pose a different challenge. How could litigants demonstrate that a program failed to protect sufficiently against a risk of error if they cannot rigorously establish the system's error rate, a measurement that needs an error standard?

BVA's QR program did not pose this difficulty, because the program failed by the terms of the agency's own error standard. Going forward, however, BVA could change the standard. It might depart from one pegged to CAVC outcomes, thereby removing an external measure of the program's success. BVA might reason, perhaps, that CAVC has an unrealistic understanding of the "reasons or bases" requirement, and that BVA will use its own threshold for what adequate reasoning is. Unless a court could establish that this error standard is flawed, litigants would struggle to establish what we do here—that the program is a failure.

466. See, e.g., *Parsons v. Ryan*, 754 F.3d 657, 676 (9th Cir. 2014).

467. See, e.g., *M.D. ex rel. Stukenberg v. Perry*, 294 F.R.D. 7, 45 (S.D. Tex.), *appeal dismissed*, 547 F. App'x 543 (5th Cir. 2013).

How best to set an error standard is a challenge that goes beyond the scope of this Article. As an initial matter, we believe that any error standard should have some connection to appellate outcomes. After all, a quality assurance program should operate as a more systematic replacement for a laborious, underutilized system of appellate review. Moreover, Congress's decision to provide for judicial review of mass adjudication represents its judgment that ultimately the federal courts have the last word on the correct administration of Social Security, immigration, and veterans' benefits. An error standard that does not include some measure pegged to what reviewing courts do with cases flouts this legislative design.

We recognize that there are good arguments to incorporate factors other than expected appellate outcomes into an error standard. An agency handling much more volume than a reviewing court, for instance, might feel obliged to provide less comprehensive reasoning than the court would prefer, lest its pace of decisionmaking slow down and worsen the backlog. Regardless, the complexity of error definition is exactly why transparency in QR and institutionalized oversight are essential, as we have argued above, and should relieve the burden on courts. Constituencies, including agency personnel as well as the affected populations, can work together to define an appropriate standard for a system of mass adjudication. External groups can then use a collaboratively crafted error standard, plus ready access to agency records, to test the success of a QR program. What an error is and how best to measure it may be difficult questions to answer in the abstract. But they should be answered, and more stakeholders than just agency officials should have a say in answering them.

(iv) *Distribution of Errors.* A final concern for structural reform litigation involves the distribution of errors. While IJ errors may disfavor immigrants, what if VLJ errors skew in favor of veterans so that an improved QR program causes awards benefits to drop? From a separation of powers perspective, the problem should give little pause. Congress establishes prerequisites for the award of veterans' benefits. If the agency systematically awards benefits to veterans who are not entitled to them, the agency usurps legislative authority. Congress may prefer generosity in the award of benefits, but systemic error in adjudication would be an ill-conceived method to vindicate this preference.

Error distribution should not create problems for a due process claim. Even if a system on balance has more false positives than false negatives, a veteran entering the system risks suffering a false negative due to poor quality assurance efforts and thus endures a risk of harm. To the veteran for whom this risk ultimately materializes, it is no answer to point to the two veterans who receive an unwarranted windfall. An immigrant denied adequate assistance of counsel in a New York City immigration court does not suffer less of a due process violation because IJs in New York overall grant relief at high rates.

In short, while the distribution of errors may explain the politics of error distribution, the distribution does not undermine the potential for structural litigation.

* * *

We have articulated theoretical implications for the study of administrative law and have spelled out the implications of our findings on disclosure, oversight, institutional design, and structural reform litigation. We make two points in closing. First, while we have sketched out policy reforms, whether such reforms are cost justified remains to be tested. There may well be diminishing returns to securing accurate decisionmaking, but such reforms have not been given serious attention precisely because existing (misleading) accuracy statistics have led Congress and stakeholders to underestimate the quality crisis in mass adjudication. Second, while our focus has primarily been on three agencies, we believe these implications are also likely to affect a much wider range of formal and informal adjudication in the administrative state. The quality of patent examinations, for example, has long been questioned,⁴⁶⁸ and the patent office has piloted a range of quality assurance initiatives,⁴⁶⁹ subject to little scrutiny. Transparency about quality improvement and rigorous evaluation of initiatives remains as vital in those areas as in the three agencies we have focused on in this Article.

Conclusion

BVA's Quality Review story does not end happily. Given the little regard many external observers have for IJ decisionmaking, EOIR's failure to implement even a rudimentary quality assurance program is shocking. SSA's recent efforts are promising, but its decades of fitful experimentation and the persistent temptation to favor quantity over quality make rigorous critique by impartial outsiders all the more essential. Prompted by the right mix of external oversight and internal incentives, however, agencies may yet vindicate *Goldberg's* commitment, not just for those individuals dogged enough to exercise procedural rights, but for everyone enmeshed in a system of mass adjudication. Two of us have helped to develop, test, and implement programs that have improved decisional quality without the commitment of significant

468. Mark A. Lemley, Essay, *Rational Ignorance at the Patent Office*, 95 NW. U. L. REV. 1495, 1495-96 (2001) (noting critiques of patent examination quality).

469. See *Office of Patent Quality Assurance*, U.S. PAT. & TRADEMARK OFF., <https://perma.cc/7D6E-4XUT> (last updated Apr. 18, 2018, 10:08 AM EDT) (listing a range of quality assurance projects).

additional resources.⁴⁷⁰ There remain plenty of veins to tap for an agency willing to experiment and subject results to external evaluation.

A due process balancing test from the 1970s will not defuse the crisis of decisional quality agencies face as they buckle under the strain of large caseloads. Internal administrative law, properly shaped by external oversight and intervention, still might.

470. See Ho, *supra* note 452, at 49-73 (describing a peer review program and its promising impact on the quality of food safety inspections in King County, Washington); *supra* notes 315-21 and accompanying text (describing BVA's short-lived experiment in 2016).