Imagine you are a rule writer in a federal regulatory agency. Suppose further that you follow developments in administrative law, knowing that the rules you write (and your subsequent interpretations of those rules in more informal settings such as guidances, policy statements, and litigation briefs) will sometimes be challenged in court as procedurally deficient, arbitrary and capricious, or in excess of statutory authority. Because of your attentiveness to matters of administrative law, you know of a doctrine—Auer (aka Seminole Rock) deference—that says that any subsequent informal interpretations and clarifications of the regulatory text you draft will be given a particularly deferential review, perhaps even more deferential than under the more well-known Chevron doctrine, which applies only to changes achieved through notice-and-comment rulemaking. Do you change your regulatory draftsmanship, writing comparatively ambiguous or opaque regulatory language in the first instance in the hope that subsequent interpretations of that text will receive a free pass?

Over the last twenty years, a growing number of scholars and even Supreme Court Justices have presumed the answer is “yes” and have proceeded from that premise to advocate jettisoning Auer deference altogether, perhaps with an eye to undermining the constitutional foundation of modern administrative law writ large. They have not hesitated to draw an inference that Auer deference sets up irresistible perverse incentives for agencies to draft underspecified rules that can be fine-tuned through interpretive rules later in the regulatory process. This practice, it is argued,

---


2 In a series of recent writings, Cass Sunstein and Adrian Vermeule have suggested that antipathy toward Auer is symptomatic of an “emerging, large-scale distrust of the administrative state, and perhaps by a belief that it is constitutionally illegitimate.” See, e.g., Cass R. Sunstein and Adrian Vermeule, The Unbearable Rightness of Auer, U. Chi. L. Rev. (forthcoming 2017), at 2; Adrian Vermeule, Law’s Abnegation: From Law’s Empire to the Administrative State 78-79 (2016) (calling the critique of Auer a “back-handed attempt to relitigate the New Deal”); Adrian Vermeule, No, 93 Tex. L. Rev. 1547 (2015) (reviewing PHILIP HAMBURGER, IS ADMINISTRATIVE LAW UNLAWFUL? (2014)); Cass R. Sunstein & Adrian, Vermeule, Libertarian Administrative Law, 82 U. Chi. L. Rev. 393 (2015). In their view, “[i]f taken seriously, general arguments from the separation of powers and general arguments about the constitutional illegitimacy of the administrative state would sweep far more broadly than the relatively modest problem of deference to agency interpretation of agency regulations.” Sunstein & Vermeule, Unbearable Rightness, supra note __, at 2. For an example of potential logical extensions, see Chris Walker’s argument that regulators’ participation in legislative drafting poses the same risk of self-delegation in the statutory interpretation as exists in regulatory interpretation, thereby raising concerns about Chevron deference. See Christopher J. Walker, Legislating in the Shadows, 165 U. Pa. L. Rev. at 4-5 (forthcoming 2017).
underscores accountability in the administrative state and erodes the rule of law by allowing agencies to move the goalpost at will, all with no public input and no meaningful judicial oversight. There have been repeated signals that several of the current Supreme Court Justices are considering overturning Auer in large part due to the logic of perverse incentives. Even supporters of Auer typically concede that the perverse incentive structure exists and that agencies frequently do avail themselves of the opportunity it allegedly presents to augment their interpretative authority.

There is just one problem: none of the Auer critics has bothered to test empirically whether rule writers do in fact systematically change their behavior in response to the alleged incentive structure. Indeed, they have not so much as offered even a single convincing anecdote consistent with the hypothesis that agencies change their rule-writing approach in response to Auer deference. It is no exaggeration to say that a good deal of Auer skepticism is based on pure

---

1 Jennifer Nou, Regulatory Textualism, 65 Duke L.J. 81, 102 (2017) (“[A]fter an agency promulgates a legislative rule through notice and comment, it can then continuously revise its interpretations without meaningful judicial notice to regulated entities and with little judicial accountability.”).

2 Matthew C. Stephenson & Miri Pogoriler, Seminole Rock's Domain, 79 Geo. Wash. L. Rev. 1449, 1453 (2011) (“[B]road judicial deference to an agency’s interpretation of its own regulations may enable an agency to enact binding rules without subjecting itself either to meaningful procedural safeguards or to rigorous judicial scrutiny.”); Kevin M. Stack, Preambles as Guidance, 84 Geo. Wash. L. Rev. 1252, 1254 (2016) (“More pointedly, critics contend that agencies rely on guidance documents in ways that circumvent the notice-and-comment rulemaking process. Their concern is that agencies are turning increasingly to guidance to establish norms that have significant de facto weight without the participation and accountability virtues of a notice-and-comment process.”).


4 See Conor Clarke, The Uneasy Case Against Auer and Seminole Rock, 33 Yale L. & Pol’y Rev. 175, 182 (2014) (“Manning’s and Scalia’s criticisms should not be exaggerated: agencies are not necessarily mechanical cost minimizers or power maximizers, and they still churn out plenty of precise, detailed rules. But these criticisms nonetheless have an attractive logical force: allowing agencies wide latitude to interpret their ambiguous regulations does seem to provide those agencies with little incentive to make their regulations more precise.”); Sunstein & Vermeule, Unbearable Rightness, supra note 2, at 13 (“In the abstract, the concern is certainly intelligible.”).

5 Ronald M. Levin, Auer and the Incentives Issue, YALE J. ON REG.: NOTICE & COMMENT (Sept. 19, 2016), http://yalejreg.com/nc/auer-and-the-incentives-issue-by-ronald-levin/ (arguing that “[a]s of yet . . . the critics of Auer have not even produced any good anecdotes to support their theory” and that the reality is “one hundred percent guesswork”); Sunstein & Vermeule, Unbearable Rightness, supra note 2, at 13 (“[W]e are unaware of, and no one has pointed to, any regulation in American history that, because of Auer, was designed vaguely and broadly.”). In fact, one of the most controversial Auer cases, Decker v. Northwest Environmental Defense Center, 133 S. Ct. 1326 (2013), almost establishes the contrapositive. Although the agency might have initially attempted to shoehorn its preferred interpretation of its Clean Water Act rules through a litigating position, in response to initial criticism it quickly pivoted strategies and promulgated its preferred interpretation through notice-and-comment rulemaking. The lack of difficulty for the agency in doing this when questioned would seem to undermine a critical assumption of the perverse-incentives hypothesis: that the notice-and-comment rulemaking process is so burdensome that it drives agencies to seek a shortcut.
speculation, untethered to any systematic evidence one way or the other. Rather than seeking to test their expectations against the record, Auer critics have instead presumed that the existence of a facially plausible mechanism establishes such extensive availment that it outweighs the many positive benefits provided by the venerable doctrine.\(^8\)

This lamentable state of affairs is no doubt attributable in part to the inherent difficulty of testing the perverse incentives hypothesis. Indeed, the late Justice Antonin Scalia—one of the foremost critics of Auer deference later in his career, despite having authored the majority opinion in Auer—noted how difficult it is to “confirm or refute this particular prediction,” in part because of uncertainty about “how one would measure [regulatory clarity] or how one would control for the various other factors that undoubtedly bear upon a regulation’s clarity.”\(^9\) But punting on the issue and still advocating for a fundamental reworking of core administrative law doctrine on little more than a hunch seems like the worst kind of judicial practice. If Auer critics want to make a persuasive case for doing away with the doctrine—an agenda still very much on the table even after Justice Scalia’s death because of the nomination of Judge Neil Gorsuch to the high Court\(^10\)—the incentives issue must be empirically explored. Relatedly, it is not enough for supporters of Auer to rely on arguments about the inevitability of balancing, as several influential accounts have. If one does not know how to measure what one is balancing—e.g., opacity versus flexibility in interstitial policymaking\(^11\)—then the exercise is meaningless, and an unfounded presumption that the devil we know is better than the devil we don’t is an arbitrary defense of the balance that Auer purportedly struck, at least so long as the perverse incentives theory looms in the picture.\(^12\)

This article provides the first systematic empirical analysis of the perverse incentives hypothesis in action. Recent advances in computerized content analysis of legal texts, including statutes and regulations, make it possible explore the impact of the Auer doctrine on the ways

---

\(^8\) Sunstein & Vermeule, Unbearable Rightness, supra note 2.


\(^10\) There is every indication that Judge Gorsuch is poised to align himself with the other remaining deference skeptics on the Court. See, e.g., Kathryn M. Schroeder & Jason B. Hutt, Gorsuch May Further Tip Balance Against Deference to EPA, LAW 360 (Feb. 14, 2017), available at https://www.law360.com/articles/890417/gorsuch-may-further-tip-balance-against-deference-to-epa; see also Gutierrez-Brizuela v. Lynch, 834 F.3d 1142, 1149 (10th Cir. 2016) (Gorsuch, J., concurring) (“There’s an elephant in the room with us today. We have studiously attempted to work our way around it and even left it unremarked. But the fact is Chevron and Brand X permit executive bureaucracies to swallow huge amounts of core judicial and legislative power and concentrate federal power in a way that seems more than a little difficult to square with the Constitution of the framers’ design. Maybe the time has come to face the behemoth.”).


\(^12\) Clarke, supra note 6, at 178. Adrian Vermeule puts more meat on the bones, noting that Auer is really about the “agency’s choice . . . to allocate its authority between more general rulemaking now and more specific interpretation or adjudication later”—a choice that is presumably itself one that the agency is best suited to make, as recognized by the Chenery. See, e.g., Vermeule, Law’s Abnegation, supra note 2, at 81 (citing SEC v. Chenery Corp., 332 U.S. 194 (1947)).
agencies write rules. In addition, the other major problem flagged by Justice Scalia—the problem of confounding explanatory variables—can be solved by employing a research design that focuses only on short windows of time when the signals from the Court to the agencies about the vitality of the doctrine are strongest. With such a strategy, it is possible to control for many of the otherwise unobservable temporal changes that might account for variation in the clarity of rules. In addition, focusing on these windows of time amounts to a most-likely case methodology: if there is indeed a systematic perverse incentives effect, we would most expect to see it in those moments when the signal is clearest. Employing this research design, I explore whether seminal Auer-related cases in the U.S. Supreme Court perceptibly caused changes in the style, clarity, and precision of rulemaking in the short term.

Part I begins with an exegesis of the debate over Auer, including the role Auer plays in administrative law, the genesis of the perverse incentives argument, and its role in the movement to overturn Auer. Part II follows by outlining in more detail the difficulties in testing the perverse incentives hypothesis, and by offering an empirical research strategy designed to tackle these difficulties. Part III presents the main results of the empirical analysis, which show little change in agency behavior in response to the changing incentives. Part IV concludes with a discussion of the implications of the findings for the debate over Auer. Although there are alternative critiques of Auer that do not rely explicitly on the unsupported perverse incentives theory, the perverse incentives argument is in practice so central to critics’ case that the null findings should stop the assault on Auer in its tracks. More generally, I use the opportunity to reflect on the appropriate methodology for interpreting regulations and to reinforce the dangers of prescribing administrative law reforms on the basis of theoretical accounts of the incentives that rules and procedures create for bureaucrats.

I. Auer’s Foundations and (Alleged) Failings

Despite the noisy debate surrounding the Auer doctrine, almost none of the controversy stems from doctrinal uncertainty. A black letter statement of the doctrine is straightforward: an agency’s interpretation of its own regulations is “controlling unless ‘plainly erroneous or inconsistent with the regulation.’” In practice, when Auer applies, it affords nearly automatic

13. Indeed, as some have noted, the other prominent critique—the separation of powers critique, see supra note 26 and accompanying text—“often blurs into the consequentialist incentives question” because the “key reason that critics dislike giving agencies legislative and interpretive power is that it creates bad incentives.” Clarke, supra note 6, at 176 n.6.


15. Auer v. Robbins, 519 U.S. 452, 461 (1997) (quoting Robertson v. Methow Valley Citizens Council, 490 U.S. 332, 359 (1989)). In recent years the Court has delineated several circumstances when Auer does not apply. See Christopher v. SmithKline Beecham Corp., 132 S. Ct. 2156, 2166 (2012) (holding that Auer is not to be afforded “when there is reason to suspect that the agency’s interpretation does not reflect the agency’s fair and considered judgment”); Thomas Jefferson Univ. v. Shalala, 512 U.S. 504, 515 (1994) (holding that interpretations that “conflict[ ] with a prior interpretation [are] entitled to considerably less deference than a consistently held agency view”); Christensen v. Harris County, 529 U.S. 576, 588 (2000)
deference, although empirical studies of agencies’ win rates when Auer is invoked in the circuit courts show that the deference is eroding, potentially in response to withering criticism from several Justices on the Supreme Court. Agency win rates under Auer in the circuit courts are now roughly on par with agency win rates under Chevron.

Auer, like many doctrines in modern administrative law, is a creature of administrative common law. The Supreme Court apparently pulled the doctrine out of thin air in 1945, just one year before passage of the Administrative Procedure Act, “cit[ing] no authority of any kind” to support the standard. Since then, the doctrine has evolved “from a restrained and unremarkable
doctrine” to a “full-blown and widely applied ‘axiom of judicial review.’”20 Yet, its creation and subsequent development should have come as no surprise. The doctrine alleviates a fundamental tension, one that would emerge in any sensible system of administrative law, and certainly in one as invested in legislative rulemaking as our administrative system—namely, “accommodating the need for agency flexibility while guarding against the specter of what Justice Jackson memorably described as ‘administrative authoritarianism’—the ‘power to decide without law.’”21 The problem Auer aims to solve, in other words, is the tension between our expectation of prior notice of the rules of the game and the inescapable need for expert application of legislative rules to more and more fine-grained, unforeseen situations that fall within the purview of the rule.22 Even the most perspicacious of rule writers cannot anticipate every circumstance that might require application of the rule.23 In theory, Auer saves the day by giving these rule writers some assurance that their good faith application of the codified text to novel, but related, problems will be respected by the courts, and it likewise gives agencies the freedom to avoid undergoing cumbersome notice-and-comment rulemaking whenever they seek to make these adjustments.24

Recent years have seen a full-throated rejection of Auer deference in the academy and the courts. At the risk of oversimplification, I suggest that critics have offered two basic critiques.25 The first sounds in concerns about the fact that “Auer allows an agency to both write the law (the regulation) and determine its application.”26 This is essentially a concern about the wisdom of the

---

21 Clarke, supra note 6, at 178 (describing these as “two classic concerns of the modern administrative state”); see also Stephenson & Pogoriler, supra note 4, at 1453 (“[W]holesale rejection of Seminole Rock would be quite disruptive, and would likely have serious disadvantages, including loss of regulatory flexibility and efficiency.”).
22 The literature identifies two somewhat related justifications of the second half of this tension (and the associated allocation of institutional authority to the agencies, rather than the courts): first, that agencies are in the best position to resolve ambiguities according to the “regulative intent,” see Stephenson & Pogoriler, supra note 4, at 1456; Sunstein & Vermeule, Unbearable Rightness, supra note 2, at 10-11, and, second, that agencies have expertise that allows them to more coherently “resolve any gaps, conflicts, or ambiguities” in the regulatory scheme, see Stephenson & Pogoriler, supra note 4, at 1456. There are important differences between these views, but I view the distinction between these justifications as unimportant for the purpose of this paper. For a general account of the unique attributes of regulatory interpretation, as opposed to statutory interpretation, see Kevin M. Stack, Interpreting Regulations, 111 Mich. L. Rev. 355 (2012).
23 Nor do they necessarily want to. See Barmore, supra note 11, at 819 (“At times, the best regulation may not be the clearest possible formulation because enhanced clarity sacrifices regulatory accessibility and congruence. An accessible rule is one that can be applied easily to concrete situations, while a congruent rule is one that produces the desired behavior. An agency could try to answer every potential interpretive question, but such clarity would increase the length and complexity of regulations until they were too opaque for regulated parties to understand.”).
24 Barmore, supra note 11, at 819-20.
25 The division maps Stephenson and Pogoriler’s identification of two “distinctive concerns that do not apply (or do not apply with equal force) in the Chevron context.” See Stephenson & Pogoriler, supra note 4, at 1460-64.
26 Barmore, supra note 11, at 817. This argument has two related derivations. Sometimes, the argument is framed in terms of a claim about the relatively greater capacity of judges to engage in the task of
current allocation of interpretive authority between agencies and courts (i.e., an inter-institutional authority optimization problem). Simply put, it reflects concern about the diminished role of courts in policing discretion in the administrative state. The second critique, which I will label the perverse incentives theory, is about the secondary effects of the allocation of authority on agency behavior—namely, “it may also have the perverse effect of undermining agencies’ incentives to adopt clear regulations.” As I discuss below, a group of Auer rehabilitators have effectively regained control of the debate over the optimal allocation of interpretive authority by fleshing out the long-missing rationale for Auer and showing that the division of institutional authority over the timing of interpretive specificity is a delicate one that admits of no unassailable answer. What has been lacking, however, is an effective answer to the perverse incentives theory. This theory, which

interpretation, both because of their expertise and because their neutrality is not compromised by having authored the very material they are interpreting. See Erica J. Shell, The Final Auer: How Weakening the Deference Doctrine May Impact Environmental Law, 45 Envtl. L. Rep. News & Analysis 10,954, 10,955 (2015) (“Enabling an agency to receive binding deference for its own interpretation in essence gives binding legal effect to that interpretation, even though “it is emphatically the province and duty of the judicial department to say what the law is.”); Orrin Hatch, The Role of Federal Judges in the Modern Administrative State, RegBlog (Mar. 27, 2017), available at https://www.regblog.org/2017/03/27/hatch-role-federal-judges-modern-administrative-state/ (“Both [Chevron and Auer] . . . displace the judiciary’s constitutional role to say what the law is.”). Other times, the argument is framed as a matter of separation of powers principle: i.e., as a formal matter, agencies violate a long-standing norm against the combination of various constitutionally prescribed powers in one body. See Sunstein & Vermeule, Unbearable Rightness, supra note 2, at 14 (“Auer’s critics have a . . . fundamental objection, one that involves heavy artillery, and that also has intuitive appeal: The decision produces a constitutionally suspect combination of the power to make law with the power to interpret law.”). In his partial dissent in Decker, Justice Scalia endorsed this perspective, writing that there could be “no congressional implication that the agency can resolve ambiguities in its own regulations” because that “would violate a fundamental principle of separation of powers—that the power to write a law and the power to interpret it cannot rest in the same hands.” Decker v. Northwest Env’tl Defense Center, 133 S. Ct. 1326, 1341 (2013). Justice Scalia sees Auer as crossing a sacrosanct line, traceable to no less than Montesquieu, Blackstone, and Publius, that “[h]e who writes a law must not adjudge its violation.” Decker, 133 S. Ct. 1326, 1342 (2013).

27 See Egan v. Delaware River Port Authority, No. 16-1471, at 2-3 (3d. Cir. Mar. 31, 2017) (Jordan, J., concurring), available at http://www2.ca3.uscourts.gov/opinarch/161471p.pdf (“As though that were not bad enough, our hand are also tied when an agency interprets or reinterprets its own rules. Those fetters were put in place by Auer v. Robbins, which extended judicial deference to an agency’s interpretation of its rules, even in the midst of litigation. The result today is that agencies are entitled to deference for their interpretation of statutes and then to a further dose of deference for their interpretation of the rules and regulations they layer on top of those statutes. All the while, federal courts are pushed further and further away from our constitutional responsibility to ‘say what the law is.’”) (internal citations omitted); Elgin Nursing & Rehab. Ctr. v. U.S. Dep’t of Health & Human Servs., 718 F.3d 488, 493 (5th Cir. 2013) (“[G]ranting deference to CMS’s interpretation of the SOM would leave no role for the courts—taken to its logical conclusion, it could effectively insulate agency action from judicial review. It is not within the province of the Executive Branch to determine the final meaning of a vague document interpreting a regulation any more than it would be to interpret the final meaning of a contract entered into by the Executive Branch.”).

28 Stephenson & Pogoriler, supra note 4, at 1461.
does not involve an optimization problem, requires a different kind of answer, and so far none has been forthcoming.

A. The Rehabilitation

Auer’s rehabilitators have responded to the critiques of the allocation of interpretive authority by doing the intellectual legwork to establish Auer’s inevitability (or at least the inevitability of some version of judicial deference to agency interpretations of regulations). As Cass Sunstein and Adrian Vermeule recently noted, such deference “greatly simplifies the task of interpreting regulation” and “reflects a sensible understanding of institutional competence.”29 Others have noted that the rule has the benefit of “providing straightforward guidance for litigation.”30 The argument for a greater judicial role seems to assume that “interpretation of ambiguities calls for purely legal skills . . . as it plainly does not.”31 Instead, “the resolution of ambiguities often involves complex issues of fact and value,” meaning that “[a]gencies, and not judges, should be settling those issues.”32

One can, of course, agree with these claims and still disagree with whether the current balance is an optimal one and whether Auer ought to be softened or changed.33 But once one accepts that the question is one of finding an optimal balance between agency flexibility and effective judicial oversight, it becomes very difficult to resist the force of Sunstein and Vermeule’s argument that “[n]othing is broken” in the current balance, or at least that there is no real way of knowing whether it is broken or why some new allocation of authority would be better at balancing the competing considerations at play.34 “The question is always relative: what’s the best way to strike the pragmatic balance?”35

Likewise, the more formalistic separation of powers critique of Auer—that it gives insufficient weight to the separation of powers values implicated by allowing an agency to engage

29 Sunstein & Vermeule, Unbearable Rightness, supra note 2.
30 Barmore, supra note 11, at 820.
31 Sunstein & Vermeule, Unbearable Rightness, supra note 2, at 3.
33 Indeed, there are numerous proposals to do so, ranging from changing the standard of review to an “intermediate type or level of deference,” i.e., Skidmore deference, see e.g., Michael P. Healy, The Past, Present and Future of Auer Deference: Mead, Form and Function in Judicial Review of Agency Interpretations of Regulations, 62 Kan. L. Rev. 633, 678 (2014); Robert A. Anthony, The Supreme Court and the APA: Sometimes They Just Don’t Get It, 10 Admin. L. Rev. 1, 11 (1996), to imposing some triggering criteria similar to Chevron’s now familiar “step zero.” See, e.g., Stephenson & Pogoriler, supra note 4, at 1484 (musing about the limitations that would make sense when the interpretation takes various forms); Clarke, supra note 6, at 189-90 (suggesting that the courts have already begun building step zero and have long had an implicit step one asking whether the rule is actually ambiguous); Sunstein & Vermeule, Unbearable Rightness, supra note 2, at 19-20 (collecting and implicitly endorsing four “checks on agencies built into the Auer framework, including “judicial enforcement of clear regulations and statutes,” “arbitrariness review,” “reliance and arbitrariness,” and “the anti-parroting canon”).
34 Sunstein & Vermeule, Unbearable Rightness, supra note 2, at 21.
35 Clarke, supra note 6, at 191.
in both “lawmaking and law-exposition”—rings hollow. According to John Manning, the apparent combination of law-making authority and interpretive authority instantiated by *Auer* constitutes a sufficiently serious breach that it justifies “a presumption that administrative schemes provide meaningful separation of lawmaking from law-exposition”—in other words, a presumption against deference to an agency’s interpretation of its own regulations. Sunstein and Vermeule rightly savage this “heavy artillery” critique. First, they note that, strictly speaking, agencies are not engaged in legislating or adjudicating when they engage in rule writing and rule interpretation, respectively—at least not in the Article I and Article III sense. Rather, they are engaged in executing the law. Relatedly, Sunstein and Vermeule note that any critique of the particular amalgam of lawmaking and law interpreting that goes into this execution of the law in the case of *Auer* would have far broader implications for the administrative state than the critics let on. Such an amalgam of institutional functions is so common in the modern administrative state that those who latch on to this critique bear a heavy burden of explaining why *Auer* must go but agencies, writ large, must stay.

One need not be entirely convinced of the “rightness” of *Auer*’s pragmatic balance to be swayed to leave *Auer* alone. If there is uncertainty, a do-nothing approach may ultimately be the only reasonable one, given that a change in the difficulty of invoking and satisfying *Auer* might lead to poorly understood changes in agency behavior in other realms. Indeed, as Aaron Nielson notes, one possible effect of overturning or watering down the *Auer* doctrine would be to discourage agencies from writing rules in the first place, instead using their prerogative under the

---

36 Manning, *supra* note 1, at 631. The discussion in *Decker* moves seamlessly from abstract political theory to the perverse incentives hypothesis discussed *infra* Part II.B, revealing that the separation of powers maxim Justice Scalia adopts is closely related to the perverse incentives hypothesis. Indeed, it is perhaps best understood as a prophylactic rule designed to ensure that no agency dares to let the incentives run wild. I suspect that this prophylactic rule has never been taken seriously because there are undoubtedly situations where intentionally incorporating ambiguity in a rule is desirable, as when an under-informed agency subject to a congressional rulemaking deadline promulgates an ambiguous rule so that it can engage in Bayesian learning. See Robert L. Glicksman & Sidney A. Shapiro, *Improving Regulation Through Incremental Adjustment*, 52 U. Kan. L. Rev. 1179 (2004). Even John Manning, who otherwise critiques *Auer*, admits that “even if it were possible to adopt pellucid and all-encompassing rules of law, the result would hardly be desirable; the appropriate level of precision in legal rules in particular contexts is a complicated issue that turns on a number of considerations.” Manning, *supra* note 1, at 647 (citing Colin S. Diver, *Optimal Precision of Administrative Rules*, 93 Yale L.J. 65, 67-76 (1983); Louis Kaplow, *Rules Versus Standards: An Economic Analysis*, 42 Duke L.J. 557 (1992)).

37 Manning, *supra* note __, at 631; see also Anthony, *supra* note 33, at 11-12.

38 Sunstein & Vermeule, *Unbearable Rightness*, *supra* note 2, at 14 (discussing the separation of powers critique—which suggests that it involves a “constitutionally suspect combination of the power to make law with the power to interpret the law”—and rejecting that critique as “both unsound and too sweeping”).


40 Id. at 16 (“If the combination of lawmaking and law-interpreting functions in agencies really is constitutionally suspect as such, then there are much larger problems than *Auer* to discuss.”).

41 Id.
Chenery doctrine to make policy via ad hoc adjudication. For critics of the balance struck by Auer—typically, those who worry more about agency lawlessness and accountability than most—this would be a decidedly worse equilibrium.

B. The Untested Perverse Incentives Hypothesis

Auer’s rehabilitators have thus answered many of the critics’ objections to the doctrine by simply showing how intractable the inter-institutional optimization problem it presents really is. But they have done little to address Auer critics’ trump card: that is, the argument that Auer causes agencies to game the system by promulgating rules that maximize their discretion to invoke deferential review of subsequent interpretations. As Justice Scalia argued in Decker v. Northwest Environmental Defense Center, if it functions in this way, Auer gives to agencies a “dangerous permission slip for the arrogation of power.” Whatever the appropriate balance of interpretive authority, widespread arbitrage of this sort would be essentially indefensible, given prevailing theories of delegation, which is precisely why this perverse incentives theory has largely survived the rehabilitation. The perverse incentives theory makes the debate a potentially nuclear one.

Hints of the perverse incentives theory emerged in Justice Thomas’s dissent in Thomas Jefferson University v. Shalala, but the filled-out perverse incentives theory is most often attributed to John Manning’s seminal article on what was then known as Seminole Rock. According to Manning, Auer has an “untoward effect upon [the agency’s] incentive to speak precisely and transparently when it promulgates regulations.” Under the doctrine, Manning argues, “if an agency issues an imprecise or vague regulation, it does so secure in the knowledge that it can insist upon an unobvious interpretation, so long as its choice is not ‘plainly erroneous.’” In its most extreme formulation, the theory is not merely that Auer eliminates an incentive to clarity, but that

43 Barmore, supra note 11, at 818 (“If Auer requires deference to an agency’s interpretation of ambiguous (but not unambiguous) regulations, agencies would maximize future flexibility and power by promulgating ambiguous regulations. An ambiguous regulation would give the agency greater discretion when deciding which enforcement actions to bring and would increase the variety of positions it could take in subsequent litigation. An agency would be free to interpret or apply the regulation in whatever (plausible) way it considers most advantageous at the time, potentially even if it has offered a different interpretation in the past. Critics worry that the incentive to promulgate vague regulations would lead to predictably more ambiguous regulations, thereby giving regulated parties less notice of prohibited or required conduct.”).
45 Thomas Jefferson Univ. v. Shalala, 512 U.S. 504, 525 (1994) (Thomas, J., dissenting) (noting that “[i]t is perfectly understandable, of course, for an agency to issue vague regulations, because to do so maximizes agency power and allows the agency greater latitude to make law through adjudication rather than through the more cumbersome rulemaking process”).
47 Manning, supra note 1, at 657 (emphasis in original).
it creates an incentive for opacity by offering the promise of aggrandizement of authority: that is, under Auer, “[a]gencies will issue vague, broad regulations, knowing full well that when the time comes, they will be able to impose the interpretation they prefer.”

Because of the public comment periods that legislative rules are typically subjected to, agencies face increasing marginal costs of precision in rule writing, but they face greatly diminished costs of precision in non-legislative rulemaking and adjudication. To make matters worse, in many instances the regulated entity only benefits from precision in legislative rulemaking; a newly precise interpretation at the stage of adjudication or litigation is typically no defense to noncompliance. In sum, the perverse incentives theory sees Auer as undermining a universally lauded features of the APA’s framework: public accountability and prior notice through rulemaking.

Although several of the Justices appear to be convinced about the perverse incentives theory and have consistently returned to the theory as they tease the possibility of overturning Auer, few commentators have recognized how much turns on actually testing the perverse incentives hypothesis. For their part, Auer rehabilitators have thought it sufficient to point out what Manning anticipated—that there are cross-cutting incentives that actually encourage

---

48 Sunstein & Vermeule, Unbearable Rightness, supra note 2, at 12.
49 Diver, supra note 36, at 73.
51 See Jason Nichols, Note, “Sorry! What the Regulation Really Means Is . . . ”: Administrative Agencies’ Ability to Alter an Existing Regulatory Landscape through Reinterpretation of Rules, 80 Tex. L. Rev. 951 (2002). But see Christopher v. SmithKline Beecham Corp., 132 S. Ct. 2156, 2188 (2012) (“It is one thing to expect regulated parties to conform their conduct to an agency's interpretations once the agency announces them; it is quite another to require regulated parties to divine the agency's interpretations in advance or else be held liable when the agency announces its interpretations for the first time in an enforcement proceeding and demands deference.”).
52 Clarke, supra note 6, at 182 (“Auer . . . suggest[s] a method for skipping the bill: agencies can comply with the barest ex ante formal procedures—say, by using notice-and-comment to promulgate a mushy legislative rule that simply restates the ambiguous text of the statute—but still give the rule bite with later interpretations that have no procedural safeguards and receive little judicial scrutiny.”).
53 Christopher, 132 S. Ct. at 2168 (endorsing the view that Auer “creates a risk that agencies will promulgate vague and open-ended regulations that they can later interpret as they see fit”).
54 But see Steve R. Johnson, Auer/Seminole Rock Deference in the Tax Court, 11 Pitt. Tax. Rev. 1, 40 & n. 213 (2013) (“[A]ware of the possibility of abuse described by Manning and Scalia, we could disregard [Auer/Seminole Rock] deference entirely because of the possibility, or we could stay our hand until a convincing record has been established that the possibility turns into actuality with sufficient frequency and consequence.”).
55 Manning, supra note 1, at 655-56 (“Of course, an agency may have various other reasons to draft clear, self-limiting rules, even when it possesses the right of self-interpretation. For instance, clear rules mean that it is less costly for regulated parties to inform themselves of the law’s requirements, and less expensive for an agency to prove noncompliance; enforcement costs will be lower when rules are transparent. Clear rules may also help an agency to exert centralized control over field officials, a factor particularly important where an agency has a large staff and handles numerous relatively small transactions. An agency may also wish to adopt clear rules to bind subsequent administrations; if an agency drafts a vague regulation leaving extensive discretion (rather than adopting a clear norm), a subsequent administration may exercise that
specificity in rule writing. These theoretical cross-cutting incentives toward clarity, which probably far outnumber the perverse incentive created by Auer, plausibly affect the behavior of both agencies and regulated entities. But there is something deeply disatisfying about this retort. Sunstein and Vermeule implicitly concede the insufficiency of this answer to the perverse incentives argument by refusing to push the cross-cutting incentives argument too far, because, as they note, “doing so would commit the sign fallacy (with a different sign).” Answering an empirical hypothesis with a theory about how other unverified empirical hypotheses interact with the original empirical hypothesis is no convincing strategy. Perhaps that is why Manning, fully aware of most of these cross-cutting incentives, nevertheless persists in the belief that the perverse incentive is stronger. There is nothing inherently unreasonable about this hypothesis about the cumulative incentives where there is no evidence about agency behavior either way. Indeed, many have noted the inherent plausibility of the theory, even as they have marshalled other arguments to rehabilitate Auer.

This is not to give the perverse incentive hypothesis a pass with an agnostic shrug. Generally, we expect people who make theoretical arguments about how the world works to carry the burden of adducing some kind of empirical evidence, whether it be qualitative or quantitative. Thus far, Auer’s critics have offered no evidence of any kind to support the perverse incentives hypothesis. As Ron Levin writes, “Auer deference is a venerable doctrine. It is deeply rooted in our history and, in its Seminole Rock formulation, older than the Administrative Procedure Act. It reflects the courts’ sound recognition of the value of the agency’s perspective in judicial review, in light of the complexity of many regulations and the agency’s responsibility for making the overall program work. . . . [T]he incentive argument is far too speculative to justify abandonment of the discretion contrary to the agency’s current preferences. If a rule is clear, the subsequent administration must incur the cost of revising the rule in order to alter the policies set by its predecessor.”

---

56 Clarke, supra note 6, at 182.
57 Levin, Auer and the Incentives Issue, supra note 7.
58 Clarke, supra note 6, at 192 (“[F]or regulated entities, Auer raises the cost of skipping notice-and-comment; Skidmore lowers the cost of skipping notice-and-comment.”); Scott Angstreich, Shoring up Chevron: A Defense of Seminole Rock Deference to Agency Regulatory Interpretations, 34 U.C. Davis L. Rev. 49, 117-18 (2000).
59 Sunstein & Vermeule, Unbearable Rightness, supra note __, at 14.
60 Levin, Auer and the Incentives Issue, supra note 7 (“When I refer to [the lack of] evidence, I do not mean to insist on specific case citations or empirical studies. As yet, however, the critics of Auer have not even produced any good anecdotes to support their theory. I have yet to read an account by a former regulator saying, ‘why, sure, I exploited the opportunities Auer creates all the time.’ Or even: ‘I remember once when I proposed a regulation, my boss responded, ‘Why bother? We can get the same deference through interpretation with much less procedural hassle.’”); id. (“[T]he factual basis of the critique of Auer isn’t paltry. It’s one hundred percent guesswork.”); Sunstein & Vermeule, Unbearable Rightness, at 14 (“There is a palpable lack of realism, and a lack of empirical grounding, to the widespread concern that Auer is a significant part of the constellation of considerations that lead agencies to speak specifically or not. We do not believe that agencies often preserve ambiguity on purpose—in fact we think that that is highly unusual—but when they do, Auer is hardly ever, and possibly never, part of the picture. The critics speak[] abstractly of possible abuses, but present no empirical evidence to substantiate their fears.”).
Others have returned to the long list of possible effects of abandoning the doctrine, including undermining "reliance interests" in "longstanding agency interpretations," destabilizing the background rules against which Congress legislates, encouraging a shift from a rulemaking-centric administrative process to an adjudication-centric one, and chilling agency efforts to offer clarification at the bequest of the regulated community.

In short, there are strong reasons to refrain from rocking the boat. The point, however, is that this cautious embrace of the "devil we know" is, in many ways, seen as insufficient by a growing chorus of Auer critics who buy a plausible, if unverified, perverse incentives hypothesis. Put simply, the "sign fallacy"—what Sunstein and Vermeule name the tendency to "identify the likely sign of an effect and then to declare victory, without examining its magnitude"—demands more than a theoretical argument in response. It demands a fair empirical analysis that estimates the magnitude, if any, of the hypothesized effect. To date, that analysis does not exist, but this article attempts to remedy that situation. The next Part details the empirical strategy to deal with what is undoubtedly a tricky problem of isolating the effect of the doctrine on a nebulous concept of "regulatory ambiguity."

II. Empirical Strategy

At the core of the perverse incentives hypothesis is a deep conceptual morass. What do proponents of the hypothesis really mean when they say that agencies write a certain way in response to Auer’s incentives? Do they mean to suggest that agencies write with less specificity and clarity across the board, or only in isolated instances or with respect to a certain kind of rule? Do they mean the resulting language is ambiguous, vague, opaque, imprecise, impenetrable, or some combination of all of these technically distinct states? Although the perverse incentives hypothesis is a paradigmatic empirical hypothesis, it is an underspecified one at best, making it necessary to make many assumptions and measurement choices if empirical testing is to take place. Nevertheless, because of its influence, the perverse incentives hypothesis needs to be articulated in a measurable and falsifiable form.

This Part attempts this articulation. While the measurement strategy and research design are not without problems, they do attempt to nail down what is of interest and subject it to scientific scrutiny. In doing so, they build on recent empirical work that attempts to measure clarity in legal texts quantitatively, and they look to the most likely cases for the perverse incentives to operate. If the results this conservative strategy yields are unconvincing, the answer is not to accept the null hypothesis—i.e., to conclude that the perverse incentives hypothesis is true. It is to demand further research with different assumptions and different measures.

61 Levin, Auer and the Incentives Issue, supra note 7.
62 Clarke, supra note 6, at 193.
63 Clarke, supra note 6, at 193.
64 Nielson, supra note 42, at __; Clarke, supra note 6, at 192.
65 Sunstein & Vermeule, Unbearable Rightness, supra note 2, at 22-23.
66 Clarke, supra note 6.
67 Sunstein & Vermeule, Unbearable Rightness, supra note 2, at 4.
A. Study Design

The perverse incentives hypothesis is about how agencies process a doctrinal signal about the robustness of the *Auer* deference regime. This signal is not static, however. At a general level, the signal has clearly weakened in recent years, as several Justices on the Supreme Court have toyed with the idea of jettisoning *Auer* altogether. At a more granular level, however, the Court’s treatment of *Auer* has oscillated between affirmation of the doctrine and questioning its validity.

For instance, the Court first hinted at possible limits on the scope of *Auer* in *Gonzales v. Oregon*, holding that the Attorney General’s interpretive rule prohibiting the use of controlled substances in carrying out otherwise state-sanctioned assisted suicides was invalid under the terms of the Controlled Substances Act. Although the Court nominally reaffirmed the validity of the *Auer* doctrine, it ultimately concluded that the doctrine was “inapplicable” because “the underlying regulation does little more than restate the terms of the statute itself.” Critically, the Court did not specify just how similar a regulation would have to be to be considered a “parroting regulation.” And the willing acceptance of the Court’s power to “simply invent[] an exception to *Auer*” undermined the stability of the doctrine, which to that point had been subject to “repeated affirmation” in the Supreme Court. Just two years later, however, the Court in *Coeur Alaska* seemed to walk back from even the anti-parroting doctrine articulated in *Gonzales*. The

---

68 The sense in which I use the word *signal* is closely related to the concept of “jurisprudential regimes” in the political science literature, in that it is based on agencies’ perceptions of a shift in the factors that determine outcomes after some critical decision by the courts. See, e.g., Xun Pang, Barry Friedman, Andrew D. Martin, & Kevin M. Quinn, *Endogenous Jurisprudential Regimes*, 20 Pol. Analysis 417, 419 (2012) (“The basic idea [of a jurisprudential regime] is that after a key precedent, we would expect to see the patterns of decision-making to be different; certain key precedents create fundamentally different doctrinal tests, such that we might expect to see an important shift in outcomes thereafter.”); Mark J. Richards & Herbert M. Kritzer, *Jurisprudential Regimes in Supreme Court Decision-Making*, 96 Am. Pol. Sci. Rev. 305, 308 (2002); Mark J. Richards, Joseph L. Smith, & Herbert M. Kritzer, *Does Chevron Matter?*, 28 L. & Pol’y 444 (2006).

69 See also Pang et al., supra note 68, at _ (discussing how regimes can feature multiple breakpoints, effectively evolving over time).

70 Healy, supra note 33, at 634-35 (“This tidy rule of deference has now come to be far less tidy and secure as a result of a quintet of recent Supreme Court decisions. These decisions—*Gonzales v. Oregon*, *Coeur Alaska, Inc. v. Southeast Alaska Conservation Council*, *Talk America, Inc. v. Michigan Bell Telephone Co.*, *Christopher v. SmithKline Beecham Corp.*, and *Decker v. Northwest Environmental Defense Center*—have had a cumulative effect on *Auer* deference that resembles the impact that *Mead* has had on *Chevron* deference.”).


72 Id. at 256-57.

73 Id. at 257.


Court noted that it would afford Auer deference to the EPA’s interpretation of its regulation—an interpretation which “essentially restate[d] the text” of the relevant statute—without even mentioning the anti-parroting principle or citing Gonzales.\(^7\) Two years later, the modulation continued when Justice Scalia penned a powerful concurrence in *Talk America* calling for repudiation of the doctrine and more or less adopting the analysis of his former clerk, John Manning.\(^7\) The signal weakened further over the next several years, with the creation of a major limitation on the applicability of Auer in *Christopher v. SmithKline Beecher*, and with Justice Scalia’s partial dissent in *Decker*—the most vociferous call for abolition of the doctrine yet—and a more modest but skeptical treatment of the precedent in Chief Justice Roberts’s concurrence.\(^7\) Finally, the Court’s center appeared to change again in 2015 in *Perez v. Mortgage Bankers Association*. While Justices Scalia, Thomas, and Alito all authored concurrences focusing on the vitality of Auer deference, Chief Justice Roberts joined Justice Sotomayor’s majority opinion, which made scant mention of the debate over Auer and appeared to downplay theoretical arguments against it. As the majority opinion put it, “[e]ven in cases when an agency’s interpretation receives Auer deference, however, it is the court that ultimately decides whether a given regulation means what the agency says.”\(^8\) This, despite the fact that much of the pre-decision commentary on the case focused on the “gathering storm surrounding Auer deference” rather than the “narrow procedural issue” presented by the case.\(^9\) Most recently, amidst a maelstrom of speculation that the Supreme Court would grant the petition for writ of certiorari in the *United Student Fund* appeal precisely to overturn Auer, the Supreme Court instead denied the petition over a boisterous dissent from the denial by Justice Thomas.\(^10\)

This oscillation in the state of the Auer regime over time provides the variation needed to conduct an empirical analysis, but it also presents certain challenges that can only be overcome by employing a quasi-experimental study design. I now outline that approach.

1. **The Regression Discontinuity Approach**

Rather than looking at all rules, or even a random sample, over a long period of time—an endeavor hopelessly complicated by historical confounders, including other determinants of rule characteristics—I focus on agencies’ proposed and finalized rules surrounding the discrete points in time where the signal has changed in such a way that attentive agencies would likely receive,
process, and adapt to the changing signal. That is, I focus on rules issued or proposed during the six months before and after several salient Auer-related Supreme Court decisions. I assume that the preceding six months reflect an equilibrium resulting from prior adjustments to the Auer deference regime. This equilibrium also reflects the whole suite of unobserved factors that determine how agencies write their rules, and these unobserved factors remain more or less constant on average even after the Supreme Court’s decision.\textsuperscript{83} The only thing that does change at each discrete break point, in fact, is the incentive structure, as determined by the direction of the signal from the court. A graphic depiction of basic intuition behind this methodology, which is a variation on a regression discontinuity design and event study analysis, is presented in Figure 1.\textsuperscript{84} Another key advantage of this methodology is that it in some ways overcomes an inherent limitation of observational studies of the perverse incentives hypothesis—that it is impossible to examine the counterfactual (i.e., what a rule would have looked like had the incentives been different). Because whether a notice is issued before or after the decision is as almost as good as random,\textsuperscript{85} it effectively approximates assignment to control or treatment group in a true experiment. The rules and proposals issued just before the intervening Auer-related decision serve as a virtual counterfactual.

\textsuperscript{83} See infra Methodological Appendix.

\textsuperscript{84} The original regression discontinuity concept envisioned analysis of individual observations clustered around an arbitrary cutoff point on a so-called “rating variable” which can be perfectly controlled for in a regression analysis. If the treatment is assigned above the cutoff point and the rating variable is controlled for, then the approach approximates an experiment. See Robin Jacob & Pei Zhu, A Practical Guide to Regression Discontinuity 1-2 (2012), available at http://mdrc.org/sites/default/files/regression_discontinuity_full.pdf; David S. Lee & Thomas Lemieux, Regression Discontinuity Design in Economics, NBER Working Paper No. 14723 (2009), available at http://www.nber.org/papers/w14723. For instance, in their seminal article, Thistlethwaite and Campbell compared students who received an award for scoring just above a threshold to those who did not receive an award because they scored just below the threshold, finding that the award meaningfully caused an increase in the chances of receiving a scholarship. See Donald L. Thistlethwaite & Donald T. Campbell, Regression Discontinuity Analysis: An Alternative to the Ex Post Facto Experiment, 51 J. Educ. Psych. 309 (1960). The regression discontinuity design has since been extended to treat time as the “rating variable,” much like an “event study” in finance. See, e.g., Lucas W. Davis, The Effects of Driving Restrictions on Air Quality in Mexico City, 116 J. Pol. Econ. 38, 48-49 (2008); Maximilian Auffhamer & Ryan Kellogg, Clearing the Air? The Effects of Gasoline Content Regulation on Air Quality, 101 Am. Econ. Rev. 2687 (2011).

\textsuperscript{85} The assumption that a subject is not privy to special information about the decision before announcement is critical. If privy to information, the subject could strategically time announcement of a rule, which would in turn eliminate the advantages of the regression discontinuity approach.
The predictions for each of the cases theorized as break points are presented below in Table 1. These predictions are based on my evaluation of what an agency lawyer surveying the legal landscape would likely think of the change to the status quo in the six-month window following the decision. However, there is no reason to discourage readers from making their own predictions. The results in Part III speak for themselves and can be interpreted in light of any theory about how each of the cases altered the Auer regime.

---

86 For instance, other analyses of the arc of cases during this period see a simple breakpoint at Talk America: before that case, Auer was on sure footing; after it, its foundation has eroded steadily. See Barmore, supra note 11, at 825-26. This narrative is somewhat consistent with the evidence presented in Part III.B.
TABLE 1: INCLUDED CASES WITH PREDICTIONS FOR THE EFFECT ON AMBIGUITY

<table>
<thead>
<tr>
<th>CASE</th>
<th>Effect on Ambiguity</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Gonzales v. Oregon</em>, 546 U.S. 243 (2006)</td>
<td>-</td>
</tr>
<tr>
<td><em>Talk America v. Michigan Bell Telephone</em>, 564 U.S. 50 (2011)</td>
<td>-</td>
</tr>
<tr>
<td><em>Christopher v. SmithKline Beecham Corp.</em>, 567 U.S. 142 (2012)</td>
<td>-</td>
</tr>
</tbody>
</table>

When *Auer* is strengthened, agencies have more reason in the immediate future to write rules that push interpretive questions to another day; but when it is weakened, agencies have greater incentives to write with clarity and precision in their initial rules.

2. Caveats to the Strategy

There are a number of natural objections to this study design. First, much depends on whether rules proposed or finalized before each seminal case are in fact similar to the rules proposed or finalized after. Indeed, the fact that the study will examine different rules at each stage may make this assumption problematic. The rules in this study concern very different subject matters, and subject matter likely determines, at least in part, how agencies write. To some extent, these differences can be controlled for by observing as many potentially explanatory characteristics as possible—e.g., whether it was a proposed rule or final rule, whether it was a major rule or not, and whether it was an original rule or an amendment to existing rules. Figures A.1 through A.3 in the Methodological Appendix present summary statistics of some of these controls. They further show that there is little variation to account for across rule type or before or after decision points. But, beyond this, these controls may not capture every relevant dimension of variation, leaving some possibility that the results are driven by chance differences in the distribution of rule texts analyzed before or after decision.

Second, the study design provides a valid estimate of the “treatment” effect of the intervening Supreme Court case only if the agencies did not anticipate the change. If they anticipated the change, then the pre-decision equilibrium (which effectively serves as a “control” for any unobserved determinants of rule text characteristics) is illusory, which in turn greatly diminishes what we can glean from this type of study. Using Supreme Court cases as “events” or “discontinuities” can be problematic in this respect—as cases almost always wend their ways through the federal judiciary for years, making most careful observers aware of what is coming down the pipeline—although that has not stopped researchers from using Supreme Court cases in
stock market event studies employing essentially this methodology.\textsuperscript{87} Even if the grant of a petition for certiorari comes as a surprise for the audience of interest, months still typically pass before the Supreme Court hears oral argument and more still before it issues its written opinion. Relatedly, the observation period of six months on each side of each case is a fairly long window compared to most event studies or regression discontinuity studies. In this case, a fairly long window was necessary, however, because of the time it can take to produce or amend a rule text. Whether the text at issue is a proposal or a finalization of a previously proposed rule, an agency that has been influenced by the Supreme Court’s decisions is unlikely to show that influence overnight.\textsuperscript{88}

Finally, it is a hallmark of regression discontinuity approaches that they have good internal validity (like experiments) but potentially poor external validity. Much depends on whether the notices analyzed match notices writ large in their characteristics. These problems can be minimized by careful and reasoned sample selection, but they are ultimately unavoidable in a regression discontinuity study.

All of these are legitimate concerns, but none of them overcomes the advantages of the approach I employ. Chief among these advantages is the fact that any other observational approach to testing the perverse incentives hypothesis that focused on continuous change over time would be even more unable to account for confounding factors that might be influencing rule writing. This problem alone would severely bias the study against finding any validity to the perverse incentives hypothesis—the null hypothesis of “no response” would be exceedingly easy to confirm amidst the noise in the data. In fact, the approach adopted here is best understood as a conservative one: a “most likely case” approach.\textsuperscript{89} If we do not see any evidence of the perverse incentives playing out here, in these moments when the signal is strongest and unanticipated, then it is unlikely we would ever be able to find empirical support for it.


\textsuperscript{88} I settled on six months as my own estimate of the amount of time it would take for an agency that was either drafting language from scratch or changing draft language in response to the case to assimilate to the new equilibrium and begin acting in accord with it, but also to not be so long that the power of the study design fades away. Even this time may not be long enough to allow agencies to adapt to the new signal from the Court, so I include a two-year window for two of the cases, \textit{Talk America} and \textit{Perez}. I selected these cases because the extended window did not overlap with a presidential election—a confounding factor, given that empirical research has shown differences in agency rule-making behavior in election years. See Anne Joseph O’Connell, \textit{Political Cycles of Rulemaking: An Empirical Portrait of the Modern Administrative State}, 94 Va. L. Rev. 889 (2008).  

\textsuperscript{89} See, e.g., Jack S. Levy, \textit{Case Studies: Types, Designs, and Logics of Inference}, 25 Conflict Mgmt. & Peace Sci. 1, 12 (2008) (“[I]f one’s priors suggest that a case is likely to fit a theory, and if the data confound our expectations, that result can be quite damaging to the theory. The inferential logic of least likely case design is based on the “Sinatra inference”—if I can make it there I can make it anywhere. The logic of most likely case design is based on the inverse Sinatra inference—if I cannot make it there, I cannot make it anywhere.”).
B. Agency Selection

The paradigmatic situation where Auer deference is implicated is when an agency issues interpretive rules (sometimes referred to as “nonlegislative rules”).\(^{90}\) These interpretive rules, which can take the form of guidance documents and policy statements, are exempted from notice-and-comment rulemaking requirements on the theory that they have no additional legal effect beyond the “parent” legislative rule, which already passed through the gauntlet of Section 553 of the APA.\(^{91}\) The most thorough empirical study to date confirms that a large number of the total Auer cases decided by the U.S. Courts of Appeals from 2011 to 2014 fell into this category of “public issuance”—a total of 50 of the 192 cases observed—and that cases involving these kinds of interpretations were among the most likely to receive deference. Fully 80 percent of the interpretations in these 50 cases received Auer deference.\(^{92}\)

What does this mean for the study design? It means that the signal is strongest when agencies go about issuing interpretive rules, such as guidances. Consequently, following the “most likely case” design, it makes sense to focus the analysis on agencies that are primed to hear and process the signal—i.e., agencies that have historically made heavy use of the two-step process of writing legislative rules and interpreting them via guidance. Fortunately, Connor Raso’s study of agency guidance practices provides some insight into which agencies fit the bill. Focusing on significant guidances and significant legislative rules, Raso computes the ratio of total guidances to total rules, with results ranging from .39 (Department of Homeland Security) to .00 (Department of Housing and Urban Development; Department of Interior; State Department; Department of Treasury; and Department of Veterans Affairs).\(^{93}\) I selected three agencies based on a combination of falling high on this scale of ratios and producing a high number of legislative rules. For instance, although the Department of Homeland Security and its predecessor agencies had the highest ratio, they produced only 170 significant rules from 1993 to 2008. Contrast this with the Department of Agriculture, which produced 1327 significant legislative rules in the same period.

I selected three agencies by optimizing the tradeoff between the size of the ratio and the denominator (i.e., the total number of significant legislative rules): the Environmental Protection Agency (EPA), which had a ratio of .19 and issued 1,095 significant legislative rules; the

\(^{90}\) Auer also applies to other kinds of interpretations, including agency adjudications and litigating positions before federal courts.

\(^{91}\) William Funk, A Primer on Nonlegislative Rules, 53 Admin. L. Rev. 1321, 1326 (2001); Robert A. Anthony, Interpretive Rules, Policy Statements, Guidances, Manuals, and the Like—Should Federal Agencies Use Them to Bind the Public?, 41 Duke L.J. 1311 (1992). It is worth pausing to note that this theory is more or less the basis for Adrian Vermeule’s “internal” critique of the Auer skeptics’ position. Citing the so-called “completion theory” of the separation of powers, which Vermeule calls the “theory of American administrative law,” he argues that the interpretations at issue in Auer are considered to be an exercise of executive authority, not legislative authority. Vermeule, Law’s Abnegation, supra note 2, at 77-78 (citing Jack L. Goldsmith & John F. Manning, The President’s Completion Power, 115 Yale L.J. 2280 (2006)).

\(^{92}\) Barmore, supra note 11, at 833.

\(^{93}\) Raso, supra note 50, at 813.
Department of Labor (DOL), which had a ratio of .20 and issued 405 significant legislative rules94; and the Department of Transportation (DOT), which had a ratio of .15 and issued 853 significant legislative rules. These agencies have in recent years promulgated some of the most high-profile and litigated rulemakings, such as the Clean Power Plan95 and the Waters of the United States96 rules at the EPA, the Overtime97 and Fiduciary98 rules at the DOL, and Corporate Average Fuel Economy Standards99 and the Rear-Camera100 rule at the DOT. On a more intangible level, these examples of high-profile rules underscore that these three agencies are subject to a critical eye, making them likely candidates to both perceive the perverse incentive structure and seek to avail themselves of it, thereby avoiding some of the scrutiny that might otherwise occur.

The very fact that these three agencies are relatively primed to respond to Auer’s incentives risks introducing problems of external validity. But, crucially, it threatens to bias the study in favor of finding an Auer effect, and likely overstates the effect on agencies that are less likely to be tuned in. To the extent possible, it is always a good idea to make potential problems with external validity into a feature, not a bug.

C. Measuring the Ambiguity of Regulatory Text

As Justice Scalia presciently observed, one of the most difficult challenges to overcome in examining the perverse incentives hypothesis is that of measurement.101 How can an abstract, situational notion of ambiguity or vagueness102 in legal texts be reliably and validly measured for a

---

94 The Department of Labor is also known for “invok[ing] Auer more often” than other agencies in litigation. See Barmore, supra note 11, at 816.
102 It is worth noting that “the concept of ambiguity is itself perniciously ambiguous.” Lawrence M. Solan, Pernicious Ambiguity in Contracts and Statutes, 79 Chi.-Kent L. Rev. 859, 859 (2004). Lawyers tend to muddle conceptual distinctions thought important by linguists and philosophers. Namely, lawyers tend to treat vagueness—i.e., the “use [of] concepts that have indefinite application to particular cases”—as a species of ambiguity, which exists when a text “can have two or more distinct meanings.” Lawrence Solum, Legal Theory Lexicon: Vagueness & Ambiguity (June 28, 2015), lsolum.typepad.com/legaltheory/2015/06/legal-theory-lexicon-vagueness-ambiguity.html. In fact, linguists and philosophers distinguish several different types of ambiguity, including syntactic and lexical ambiguity, and they sharply distinguish vagueness from these formal types of ambiguity. See, e.g., Maryellen MacDonald, The Interaction of Lexical and Syntactic Ambiguity, 32 J. of Memory & Lang. 692 (2001). Rather than wedding the study to any technical notion of ambiguity or vagueness, the measures that follow in Part II.C.1-4 follow the lawyer’s intuition, treating
large enough sample of texts to support statistical analysis? The short answer is that it cannot, at least not directly.103

But what can be done—and indeed what has been done in a growing number of empirical legal studies—is to rely on a defensible measure of the potential for semantic slippage to arise through opacity. In fact, when it comes to testing the perverse incentives hypothesis, we are less interested in any actual identifiable ambiguity than we are with measuring this potentiality, born of chronic underspecification. After all, the agency availing itself of the perverse incentive structure may have only a general idea of what concrete actions it hopes to take in future iterations of policymaking; what it really wants to accomplish is to maximize flexibility and leave ample opportunities to flesh out its agenda with less constraint. While deliberately leaving one key term underspecified in hopes of accomplishing a specific purpose in the future—a kind of reverse “elephant in a mousehole” tactic104—is a plausible way that the perverse incentives could manifest, it is also isolated and of less practical concern. The critique of Auer only has any power at all because it sweeps far more broadly, suggesting that agencies systematically underspecify the rules they make.

That still leaves the problem of deciding on what this proxy measure of opacity should look like and how it can be implemented. Here, I review the merits of four widely used proxy measures for opacity, or lack of clarity and specificity, in legal texts.

1. Length of Document

A number of studies in political science have relied on the length of statutes (texts which are in many ways similar to regulations105) to measure latent discretion, which could be seen as closely related to ambiguity. The theory is that “the more words a legislature puts into legislation on the same issue, the more it constrains other actors who will implement policy on that issue,” and “the fewer words it writes, the more discretion it gives to other actors.”106 For instance, in one line of studies, Kirk Randazzo, Richard Waterman, and colleagues employ simple word counts of the statutes at issue in cases in state and federal courts and show that ideological voting is tempered when the statute is relatively prolix.107
While the word-count measure in these studies is often extensively tested for validity\textsuperscript{108} and seems well suited to capture the discretion given to a delegate in a principal-agent relationship,\textsuperscript{109} it may be that the measure is orthogonal to the concepts of clarity or specificity. As David Law and David Zaring observe, “the text of a short statute may offer judges very little in the way of guidance, and “longer statutes contain, in a literal sense, more room for ambiguity and inconsistency.”\textsuperscript{110} For this and other reasons, “it is not entirely clear whether one should expect [statutory bulk’s] overall effect to be positive or negative” in terms of clarity.\textsuperscript{111}

Nevertheless, because it is a commonly included proxy for a lack of clarity, I include a measure of length in this study. Here, I measure the length of each regulatory text by taking the natural log of the inverse of the total word count for each document. While I acknowledge the interpretive dilemma flagged by Law and Zaring, I operationalize the variable to reflect the theory that the shorter the text, the less guidance there is, and therefore the less clear it is. Conversely, the longer a statute is, the more specific it is.\textsuperscript{112} This, it seems, is more consonant with what \textit{Auer} critics focus on when articulating the perverse incentives theory. Take the anti-parroting rule announced in Gonzales, which was designed to snuff out underspecification by forcing agencies to write more.\textsuperscript{113} The theory here appears to be that filling out the statutory command with more detail, nuances, and explication increases precision and works out lingering ambiguities.

2. Readability

A more intuitively plausible approach to measuring ease of interpretation is to measure the readability of statutes. Readability captures well the concept of complexity, and where something is


\textsuperscript{108} Huber and Shipan’s study is exemplary in this regard. They center on word count as a reasonable proxy only after a “qualitative and quantitative investigation of a huge number of statutes suggest[ed] that the more words a legislature puts into legislation on the same issue, the more it constrains other actors who will implement policy on that issue.” Huber & Shipan, \textit{supra} note 106, at 73.

\textsuperscript{109} After all, the length of a statute is itself a signal from the principal to its agent (and to other actors that help police the delegation) about the principal’s seriousness about constraining the agent. Wholly apart from the actual words used by the statute, we would expect the statute’s length to be a useful heuristic to courts engaged in reviewing agency action, which may in part explain the findings in Huber & Shipan, \textit{supra} note 106, and the sources cited in footnote 107, \textit{supra}.

\textsuperscript{110} David S. Law & David Zaring, \textit{Law Versus Ideology: The Supreme Court and the Use of Legislative History}, 51 Wm. & Mary L. Rev. 1653, 1690-91 (2010).

\textsuperscript{111} \textit{Id}.


complex, it is more likely, on the whole, that there is room for interpretation. There are at least twenty widely used measures of readability, ranging from the widely known Flesch-Kincaid scores to more obscure items, such as the Gunning Frequency of Gobbledygook (FOG) Index. Generally speaking, each of these measures uses content analysis to measure the “words, sentences, characters, and syllables” in a given text and produce indices generated from various combinations of these factors. While each of these measures is slightly different, they are all more or less geared to provide “quantitative, objective estimates of the difficulty of reading selected prose.”

Here, I elect to use the Flesch-Kincaid Grade Level measure, which follows the formula (.39 X (Total Words/Total Sentences) + 11.8 X (Total Syllables/Total Words) – 15.59), using the open source program Flesh to compute the scores. The scores are relatively simple to interpret: the grade-level score theoretically reflects the years of schooling necessary to easily comprehend the text.

Although readability scores were not developed with legal texts in mind, increasingly they have been deployed to capture the clarity of such texts. Empirical legal scholars employing readability measures in the non-statutory context have also shown that readability predicts actual comprehension and retention of information when reading Supreme Court cases, human coders’ own subjective interpretations of the clarity of the text, and the subjective predictions about the likelihood that the text would compel compliance from those subjected to the opinion. Readability measures have also been successfully deployed to gauge readability of Supreme Court opinions and to explain when and why Supreme Court Justices might deliberately (or subconsciously) author more or less readable opinions.

What about the validity of the measure for statutes? Only one study used readability measures to gauge clarity of statutes: Law and Zaring’s study of the factors that determine when courts resort to legislative history to interpret statutes. The results, while not a definitive validity test, do seem intuitively plausible as an explanation of the underlying construct of clarity. Among the least-clear statutes in Law and Zaring’s dataset are the Robinson-Patman Antitrust Act, the Voting Rights Act, the Federal Tort Claims Act, the Internal Revenue Code, and the Longshore & Harbor Workers’ Compensation Act, all of which will register with most trained lawyers as about

---

114 Cf. Law & Zaring, supra note 110, at 1691 (noting that statutory complexity has a clear effect on the likelihood of a court resorting to legislative history, and using the Flesch-Kincaid Grade Level measure to capture the complexity of federal statutes).
118 Black, Owens, Wedeking, & Wohlfarth, supra note 115, at 53-59.
119 See, e.g. id.; Ryan J. Owens, Justin P. Wedeking, & Patrick C. Wohlfarth, How the Supreme Court Alters Opinion Language to Evade Congressional Review, 1 J. L. & Courts 35 (2013) (finding that ideologically distant Supreme Court panels author less readable opinions, ostensibly to increase the costs of congressional monitoring and override).
And while readability measures “have their limitations as actual measures of the readability of legal texts, particularly at extreme levels of linguistic complexity that they were not intended to capture,” it bears noting that the rule texts and preambles analyzed in this study are far less extreme and far more consistent than the federal statutes analyzed by Law and Zaring. The mean rule text in the study is written at the level of a sophomore in college, and the mean preamble is written at the level of a first-year graduate student. At these levels, the Flesch-Kincaid Grade Level measure can be expected to suitably measure the complexity of the texts.

**Figure 2: Comparing Flesch-Kincaid Grade Levels of Rule Texts, Preambles, and Federal Statutes**

Notes: The scores for federal statutes are those reported by Law & Zaring, supra note __, at 1720.

In short, readability statistics are a well-validated method of determining just how accessible a legal text is, and thus a poor readability score is a reasonable proxy for the kind of writing that Auer critics mean to attack.

3. Cognitive Complexity

The way we write reflects the structure of our thinking: “[W]ords are the very stuff of psychology and communication.” Increasingly, empirical legal scholars analyze the types of

---

120 Law & Zaring, supra note 110, at 1700-01. The study also found that “complexity,” as defined by a high Flesch Kincaid Grade Level score for the statute, predicted the use of legislative history at a statistically significant level. See id. at 1720.

121 See Figure 2, infra.

words used in legal texts to probe the cognitive complexity displayed by the author—a marker which can in turn serve as a “proxy for clarity.” This follows because high cognitive complexity reveals a tendency to “interpret events in multidimensional terms and to integrate a variety of evidence in arriving at decisions” and low cognitive complexity denotes “conceptual organization of decision relevant information.” More specifically, cognitive complexity consists of two related concepts called “differentiation” and “integration.” The former “represents the degree to which an individual can see multiple perspectives or dimensions in an issue,” and the latter indicates “the degree to which a person recognizes relationships and connections among these perspectives or dimensions and integrates them into their decision or judgment.” There are obviously important parallels with the literature on rules versus standards. For example, Owens and Wedeking find that Justice Scalia, a well-known enthusiast of rules, wrote opinions that displayed comparatively low scores on measures of cognitive complexity.

I theorize that cognitive complexity is well suited to proxy for the kind of underspecification that critics of Auer see as a likely result of the perverse incentive structure. When an agency writes in a cognitively complex style, it is in effect leaving itself room to emphasize different facts or different considerations in the next round of decisionmaking. For instance, Owens and Wedeking find that measures of cognitive complexity in the opinions authored by Supreme Court justices predict ideological drift over the course of their careers. This is precisely what critics of Auer see agencies doing—i.e., using underspecification in the first round of play to arbitrage for greater authority or discretion in future situations. In other words, cognitive complexity proxies for the capacity for plausible drift.

Much as with readability scores, recent developments in computerized content analysis have made the task of measuring cognitive complexity much less resource intensive and far more reliable than human coding. Following the lead of a number of recent studies of Supreme Court opinion writing, I use the program Linguistic Inquiry and Word Count (LIWC) to measure several linguistic dimensions of text that combine to measure the core attributes of cognitive complexity.

123 Black, Owens, Wedeking, & Wohlfarth, supra note 115, at 44.
126 Black, Owens, Wedeking, & Wohlfarth, supra note 115, at 44.
complexity, differentiation and integration. As a general matter, LIWC proceeds from the general observation that function words, as opposed to content words, are often the key to understanding the underlying cognitive structure of a piece of text.\textsuperscript{131} LIWC searches a text and identifies what percentage of the function words (and associated word stems) in the text fit into various dictionaries designed to capture cognitive dimensions of the writing.\textsuperscript{132} Prior work on cognitive complexity and legal writing has employed several of LIWC’s built-in dictionaries to construct a general index.\textsuperscript{133} Although several of LIWC’s dictionaries have been changed, replaced, or augmented since this earlier work, the core components of the index are still available. These include causation words,\textsuperscript{134} insight words,\textsuperscript{135} discrepancy words,\textsuperscript{136} tentativeness words,\textsuperscript{137} certainty words,\textsuperscript{138} difference words,\textsuperscript{139} negation words,\textsuperscript{140} and six-letter words.\textsuperscript{141}


\textsuperscript{133} Owens & Wedeking, \textit{Justices and Legal Clarity}, supra note 128.

\textsuperscript{134} These are words that denote a causal relationship between ideas, concepts, or items. Some of the words in this dictionary, which includes 135 words, are because, effect, hence, and depend. Higher scores indicate greater cognitive complexity. \textit{See} Owens & Wedeking, \textit{Justices and Legal Clarity}, \textit{supra} note 128, at 1055-56.

\textsuperscript{135} These are words that denote depth of understanding and layers of problems. Some of the words in this dictionary, which includes 259 words, are think, know, and consider. Higher scores indicate greater cognitive complexity. \textit{See} Owens & Wedeking, \textit{supra} note 128, at 1056.

\textsuperscript{136} These are words that denote recognition of tension or inconsistencies between two or more ideas. Some of the words in this dictionary, which includes 83 words, are should, would, and could. Higher scores indicate greater cognitive complexity. \textit{See} Owens & Wedeking, \textit{supra} note 128, at 1056.

\textsuperscript{137} These are words that denote hesitation. Some of the words in this dictionary, which includes 178 words, are maybe, perhaps, and fairly. Higher scores indicate greater cognitive complexity. \textit{See} Owens & Wedeking, \textit{supra} note 128, at 1056.

\textsuperscript{138} These are words that denote confidence about a position or assertion. Some of the words in this dictionary, which includes 113 words, are always, never, and absolutely. Higher scores indicate lower cognitive complexity. \textit{See} Owens & Wedeking, \textit{supra} note 128, at 1056.

\textsuperscript{139} The difference words dictionary replaced and bolstered the 2007 dictionary for exclusive words. See James W. Pennebaker et al., \textit{Linguistic Inquiry and Word Count: Operator’s Manual} (2015). These are words that denote a disconnect or differentiation between ideas or things. Some of the words in this dictionary, which includes 81 words, are hasn’t, but, and else. Higher scores indicate greater cognitive complexity. \textit{See} Owens & Wedeking, \textit{Justices and Legal Clarity}, \textit{supra} note 128, at 1056.

\textsuperscript{140} These are words that denote the contrapositive of an affirmative statement or position. Some of the words in this dictionary, which includes 62 words, are no, not, and never. Higher scores indicate lower cognitive complexity. \textit{See} Owens & Wedeking, \textit{Justices and Legal Clarity}, \textit{supra} note 128, at 1056.

\textsuperscript{141} As the name suggests, these are words that contain over six letters. They are thought to indicate sophistication and complexity. \textit{See} Owens & Wedeking, \textit{Justices and Legal Clarity}, \textit{supra} note 128, at 1056.
I adopt this approach, generating a summary measure—Cognitive Complexity—that captures the relative complexity of the texts in the dataset. After obtaining raw percentage scores for each of the categories theorized to relate to cognitive complexity for each text, I standardize the scores by obtaining z-scores for each category in each text. Finally, again following the method outlined in prior studies, I use the standardized scores to create an aggregate measure of cognitive complexity for each text in the dataset.

4. Linguistic Obfuscation

A final proxy measure focuses more explicitly on key markers in language that indicate intent to deceive through linguistic obfuscation. Research into fraudulent scientific research finds that writers who intend to deceive do so not only by decreasing the readability of their writing, but also by utilizing “higher rates of causal terms (e.g., depend, infer),” “more abstraction (e.g., fewer articles, prepositions, and quantifiers),” and “more jargon than nonobfuscated language.” While even the harshest critics of Auer would be unlikely to say that an agency is literally lying when it writes an ambiguous regulation in response to perverse incentives, there are important and obvious parallels between the kind of deceptive writing documented by this line of linguistic research and the kind of obfuscation of purpose that agencies are alleged to engage in with Auer. In both cases, the writer is trying to increase the costs of analyzing the true meaning (or non-meaning) of the text by resorting to predictable linguistic patterns.

Following this line of research, I use LIWC to create a measure of linguistic obfuscation for the regulatory texts in the dataset. The measure adds together two standardized textual measures. First, it accounts for “jargon” by determining how much of the text is composed of “specialized terms that are uncommon in everyday communication.” Using LIWC makes this measurement easy: any word that is not contained in the large dictionary is considered specialized jargon. Following Markowitz and Hancock’s study of scientific deception, I subtract the percentage of words identified by LIWC from 100 and then generate z-scores to standardize the results. Second, it accounts for abstraction by “taking the inverse of the sum of the standardized LIWC scores for

\[ z = \frac{x - \mu}{\sigma} \]

where x is the observed raw score, \( \mu \) is the mean score for all texts, and \( \sigma \) is the standard deviation.

---

142 The z-score is simply \( z = \frac{x - \mu}{\sigma} \), where \( x \) is the observed raw score, \( \mu \) is the mean score for all texts, and \( \sigma \) is the standard deviation.

143 See Owens & Wedeking, Justices and Legal Clarity, supra note 128.

144 Id.

145 David M. Markowitz & Jeffrey T. Hancock, Linguistic Obfuscation in Fraudulent Science, 35 J. Lang. & Soc. Psycy. 435, 436 (2015). In this sense, the literature on fraudulent research underscores the plausibility of simple readability measures as a marker of certainty. See infra Part III.C.2.

146 Markowitz & Hancock, supra note 145, at 436.

147 Markowitz & Hancock, supra note 145, at 441 (“The obfuscation literature suggests that fraudulent financial companies dissimulate their reports to make them more costly to analyze.”).

148 Markowitz & Hancock, supra note 145, at 438.

149 Pennebaker et al., supra note 132; Tausczik & Pennebaker, supra note 122.
articles, prepositions, and quantifiers.” When the score on this variable is high, it “suggests that the language is less descriptive and less concrete than [when there is] a low abstraction score.”

D. Putting It All Together

To briefly recap, this Part outlined a measurement strategy and study design which will be used in Part III to test the perverse incentives hypothesis: that is, the hypothesis that agencies respond to Auer deference by systematically withholding specificity or obfuscating their purpose when they write legislative rules in the hopes that they will be able to maximize their discretion to later make interpretive changes and adjustments without rigorous oversight and without affording a robust opportunity for oversight. At the core of this hypothesis is the idea that agencies can dial back or ratchet up the clarity of the rules they promulgate in response to probability that the agency’s interpretation in guidance documents, policy statements, litigation briefs, and the like will be given Auer deference. As agencies see Auer eroding, they are less likely to employ this strategy, as the courts are more likely to subject the interpretation to scrutiny that will uncover the weak link between the parent rule and the subsequent interpretation; as agencies see Auer strengthening, they are more likely to try to take advantage of the leeway courts give any plausible interpretation. If the perverse incentives hypothesis is to have any bite, it must take something like this systematic form.

If this formulation of the hypothesis is correct, then the appropriate strategy to test the hypothesis is to focus on those times when the health of the Auer regime is in flux—i.e., at times when the Supreme Court rules on Auer and changes the expected payoff for an agency to delay interpretive specification. By ironing out our expectations as to the direction of the signal sent from the Court, we can examine the rule-writing behavior (as captured by the four measures of regulatory clarity discussed in Part II.C) of three of the agencies most primed to respond to Auer’s incentives, both immediately before and immediately after the decision. Any observable trend or difference in this window of time that coincides with the predicted direction of the change might be plausibly attributed to Auer’s incentives.

Thus, in Part III I estimate the effect of changes in the Auer regime using two strategies, both of which are based on the inferential logic discussed in Part II.A. First, following work on the effectiveness of pollution-control policies, I estimate generalized least squares (GLS)
regressions of measures of the proxy measures of ambiguity on whether the text was written before or after the Auer-related decision.\textsuperscript{154} Second, I use regression discontinuity (RD) methods designed to capture the instantaneous treatment effect at the point of decision. The basic idea is to use localized regression within “bins” of observations immediately before and after the treatment, providing what is in effect a weighted average that fits the data in that bin, and then subtracting the average immediately before the treatment from the average immediately after treatment.\textsuperscript{155} Unlike the GLS models, these RD models allow for localized estimates that provide a more fine-grained look at the immediate fluctuations around the decision point.

The approach I have adopted is helpful for many reasons. Focusing on narrow windows of time around a decision point helps to control for the many other unrelated influences on agency rule writing, on the theory that nothing fundamental besides the signal should have changed in the interim.\textsuperscript{156} Over a long period of time many factors conspire to influence the clarity of regulations, including the nature and specificity of the source of the regulations—i.e., statutes.\textsuperscript{157} For one example, as statutes layer multiple and overlapping responsibilities on agencies,\textsuperscript{158} the task of writing clear rules most likely becomes more difficult. Other determinants of specificity are embedded in the very task of writing effective rules, which involves trading off “transparency,” “accessibility,” and “congruence,”\textsuperscript{159} but these too can change over time. Parsing out the relative influence of these kinds of changes versus any changes attributable to a change in Auer deference over a long period time would be next to impossible. But within the short windows of time analyzed in this study, we can plausibly assume that the outside determinants of rule ambiguity remain, on average, constant, leaving any observable change plausibly attributable to the intervening change to the doctrine. In other words, because these unobserved confounders are

\textsuperscript{154} Formally, I estimate \( y_t = \gamma_0 + \gamma_1 1(\text{prior/sub}) + \gamma_2 x_t + u_t \), where \( y_t \) is one of the four measures of text clarity, \( \gamma_1 \) is the coefficient for the indicator variable prior/subsequent, which measures whether the text observed was published in the Federal Register before or after the decision by the Court, and \( \gamma_2 \) represents the coefficients for a vector of covariates, \( x_t \), describing basic characteristics each rule, such as the level of significance and whether it is expected to affect small entities.

\textsuperscript{155} More formally, the regression discontinuity treatment effect in a “sharp” regression discontinuity design is expressed as follows: \( \tau = \tau(x) = \mathbb{E}[Y_i(1) - Y_i(0)|X_i = x] \), where \( Y_i(1) \) is the regression function immediately after the decision and \( Y_i(0) \) is the regression function immediately before the decision. See Sebastian Calonico et al., \textit{rdrobust: Software for Regression Discontinuity Designs}, Stata Journal, at 5 (forthcoming 2017), available at http://www-personal.umich.edu/~cattaneo/papers/Calonico-Cattaneo-Farrell-Titiunik_2017_Stata.pdf.

\textsuperscript{156} Davis, \textit{supra} note 153, at 49 (noting that “confounding factors” can be controlled for by “considering an arbitrarily narrow window of time around the implementation of [the indicator variable for the policy treatment]” because, “[w]ithin this interval, the unobserved factors influencing [the dependent variable] are likely to be similar so that observations before [the treatment] provide a comparison group for observations after [the treatment]”).


\textsuperscript{158} Jacob E. Gersen, \textit{Overlapping and Underlapping Jurisdiction in Administrative Law}, 2006 Sup. Ct. Rev. 201, 203 (arguing that “overlapping and underlapping jurisdictional assignment can produce desirable incentives for administrative agencies” and that they can serve as “useful tools for managing principal-agent problems inherent in delegation”).

\textsuperscript{159} Diver, \textit{supra} note 36, at 67.
slow-moving, focusing on snapshots of time effectively controls for them. As long as there are no unobserved discontinuities, or spikes, in the immediate vicinity of the decisions, the two methods I employ should produce unbiased estimates of the treatment effect.

Relatedly, the strategy has the benefit of focusing on those times when the signal is loudest. Auer deference is not well known among agency rule writers, and, under normal conditions, even those who are aware of it are not actively thinking about it. But when the Supreme Court speaks, lawyers tend to listen, and we would expect agency rule writers to be especially primed to receive the signal and incorporate it into their work. In other words, the immediate aftermath of an Auer-related decision is, more than any other time, when we would expect Auer’s incentives to play out as expected. In this way, the strategy amounts to a conservative, most-likely case for detecting a response to Auer’s incentives.

III. Empirical Results

As we have seen, there is a widespread and intuitively plausible theory that the level of deference courts afford agency interpretations of their own regulations influences the way agencies write their regulations. But as we have also seen, rule writers are exposed to cross-cutting incentives, such as the desire to maximize compliance by regulated entities and to minimize enforcement costs, that otherwise tend to muddy the waters. So far, the debate over Auer has been waged on a theoretical plane, with different assessments of the ubiquity and force of the competing incentive structures largely governing whether one is for or against Auer reform. While the approach I have outlined above in Part II may have identifiable imperfections—it is hard to imagine what a “perfect” approach might even look like—it at least promises to rise above this parlor game and ground our consideration of Auer’s incentives in systematic evidence. The design aims to filter out the effect of the cross-cutting incentives, isolating only the effect of changes in the Auer regime as an explanation for any observed changes in agency behavior.

As detailed below, the simple takeaway from the analysis is that changes to the Auer deference regime have little systematic effect on agencies’ rule writing. In select situations, the changes do have effects that are clear, but they are in most cases systematically at odds with the predictions the perverse incentives hypothesis would offer—indeed, they suggest that doing away with Auer could push agencies to compensate for the loss of deference by obscuring their rules to the point that the court would be unlikely to effectively police interpretations. And while the analysis does not speak to whether the agencies respond to changes in Auer deference by increasing their reliance on interpretive rules—the second link in the perverse incentives chain—it does show that agencies are not artificially increasing their ability to do so by promulgating more ambiguous rules in the first instance. This is not to say that the perverse incentives do not exist; it is simply to say that they do not appear to visibly manifest themselves in changed agency behavior, which is really all that should matter for the normative and legal debates over the deference doctrine.

160 Walker, supra note 157, at 1065-66 (reporting on interviews with agency rule writers).
A. Global Analysis

As a first cut, I analyze all of the Auer-weakening cases together and all of the Auer-strengthening cases together. Because each window extends backwards for six months and forward for six months from a single point in time, I am able to analyze each of the cases with the same hypothesized direction of changed incentives together. To be clear, this level of analysis may obscure some of the nuances in the data. Even if the general direction of the signal is fairly identifiable, the intensity of that signal is more difficult to predict. More problematic yet, if even one of the cases does not display an effect while the others do, the effect might disappear in the aggregate results. This is why Part III.B breaks down the analysis in gory detail, case by case. But, on the whole, a global effect is what Auer critics have hypothesized, so that is where we start.

Table 2 shows the main results of the global analysis of Auer-strengthening decisions on ambiguity. That is, it shows the results for the windows of observation around the four cases that reinforced the vitality of Auer deference: first, Auer itself, followed by Coeur Alaska, and finally by Perez and United Student Aid Fund. The table includes two estimates: a generalized least squares regression estimate of the effect of the dummy variable prior/subsequent on each of the proxies for ambiguity discussed in Part II.C, and a regression discontinuity estimate that captures the local linear effect within the bandwidth specified, measured in days. For instance, the first discontinuity estimate in Table 2 shows a negative effect of .024 on the measure of wordcount for the 55.82 days after the decision, compared with a substantially smaller negative effect of .003 for the linear estimate for the entirety of the period after the decision. The bandwidth for the regression discontinuity estimates is automatically optimized based on the distribution of the data. I break the analysis out by two main categories: first, whether the text analyzed was the actual rule text attached to the Federal Register notice or instead the “preamble,” and, second, whether the text observed was part of a final rule or proposed rule. One might suspect that proposed rules would be more likely to respond to changes in Auer’s incentive structure in the short term, as significant changes to a final rule would be limited by the logical outgrowth rule. One might also suspect

161 The GLS approach equally weights observations occurring at the tail ends of the observation window with those occurring earlier in the observation window and assumes a simple linear relationship between the forcing, or rating, variable and the outcome variable. Christopher Skovron & Rocio Titiunik, A Practical Guide to Regression Discontinuity Designs in Political Science, 13 (2015), available at http://www-personal.umich.edu/~titiunik/papers/SkovronTitiunik2015.pdf; Andrew Gelman & Guido Imbens, Why High-Order Polynomials Should Not Be Used in Regression Discontinuity Designs, NBER Working Paper No. 20405 (2014). In other words, the GLS models are essentially a regression discontinuity design using the full six-month period on each side of the decision as the bandwidth for estimation, which makes them well-suited for capturing the longer-term change post-decision.

162 The regression discontinuity estimates included here capture more ephemeral changes immediately surrounding the “treatment”—specifically, they capture the difference in the local-linear estimate in a short-term bandwidth surrounding the treatment. By their very nature, these kinds of models are less-suited to capture slow-moving change.

163 See Phillip M. Kannan, The Logical Outgrowth Doctrine in Rulemaking, 48 Admin. L. Rev. 213, 214-15 (1996) (“In its broadest form, the logical outgrowth test requires that the final rule has been presaged adequately in the notice of proposed rulemaking.”); see also Cass R. Sunstein, The Office of Information and
that preambles are where much of the action is likely to be found, as it is there that the agency outlines its purposes and reasoning—information which can become critical to a court as it reviews any subsequent interpretation, just as much so as (if not more so than) it does with legislative history in the statutory context.\textsuperscript{164}

The first notable result in Table 2 is the lack of any statistically significant result. None of the specifications rises to the level that the observed effect can be safely distinguished from a true effect of zero. This can be seen most clearly in Figures 3 and 4, which show the linear estimates for the prior/subsequent variable for the full six-month periods before and after decision. With each proxy for ambiguity, the observed jumps are small and well within the bounds of measurement error. To be clear, this does not prove the null hypothesis that there is no effect: it only proves that the signal is just too weak to rise above the noise. But that is, in itself, an important finding.

Second, even were these results significant, they would often fail to accord with the theory-based predictions in Table 1 about the direction of the change post-decision. Theoretically, the preponderance of the signs on the estimates in Table 2 should be positive. And with respect to preambles, there is indeed slightly more consonance with theoretical expectations than there is with respect to rule texts. But overall there is no discernible pattern in the data, and certainly not one consistent with the idea that agencies respond to a strong Auer deference regime by being more cryptic than they otherwise would be. While it would perhaps be too much to expect every instance to comport with theory, one would expect more consistency and strength in these results before concluding that Auer-strengthening decisions lead to more opacity.

\textit{Regulatory Affairs: Myths and Realities}, 126 Harv. L. Rev. 1838, 1870 (2013) (discussing the logical outgrowth rule and noting that “a great deal of the OIRA process can be devoted to the issue”).

\textsuperscript{164} Stack, \textit{Interpreting Regulations}, supra note 22.
Table 2: Estimates of the Effect of Auer-Strengthening Decisions on Proxies for Ambiguity

<table>
<thead>
<tr>
<th></th>
<th>Final Rules</th>
<th>Proposed Rules</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Rule Text</td>
<td>Preamble</td>
<td>Rule Text</td>
</tr>
<tr>
<td>Word Count</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linear Estimate</td>
<td>-0.003</td>
<td>0.001</td>
<td>-0.001</td>
</tr>
<tr>
<td></td>
<td>p 0.552</td>
<td>0.602</td>
<td>0.883</td>
</tr>
<tr>
<td></td>
<td>R² 0.196</td>
<td>0.272</td>
<td>0.122</td>
</tr>
<tr>
<td></td>
<td>p&gt;χ² 0.000</td>
<td>0.000</td>
<td>0.211</td>
</tr>
<tr>
<td>Discontinuity</td>
<td>-0.024</td>
<td>0.000</td>
<td>0.025</td>
</tr>
<tr>
<td></td>
<td>p 0.229</td>
<td>0.960</td>
<td>0.244</td>
</tr>
<tr>
<td></td>
<td>bw 55.820</td>
<td>45.370</td>
<td>35.520</td>
</tr>
<tr>
<td>N</td>
<td>200/202</td>
<td>201/203</td>
<td>144/145</td>
</tr>
<tr>
<td>F-K Grade Level</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linear Estimate</td>
<td>0.544</td>
<td>-0.213</td>
<td>-0.063</td>
</tr>
<tr>
<td></td>
<td>p 0.234</td>
<td>0.303</td>
<td>0.904</td>
</tr>
<tr>
<td></td>
<td>R² 0.172</td>
<td>0.101</td>
<td>0.100</td>
</tr>
<tr>
<td></td>
<td>p&gt;χ² 0.000</td>
<td>0.108</td>
<td>0.429</td>
</tr>
<tr>
<td>Discontinuity</td>
<td>1.710</td>
<td>0.849</td>
<td>-1.600</td>
</tr>
<tr>
<td></td>
<td>p 0.295</td>
<td>0.298</td>
<td>0.482</td>
</tr>
<tr>
<td></td>
<td>bw 50.080</td>
<td>48.390</td>
<td>44.740</td>
</tr>
<tr>
<td>N</td>
<td>201/203</td>
<td>201/203</td>
<td>144/145</td>
</tr>
<tr>
<td>Cognitive Complexity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linear Estimate</td>
<td>-0.084</td>
<td>0.463</td>
<td>-0.165</td>
</tr>
<tr>
<td></td>
<td>p 0.884</td>
<td>0.383</td>
<td>0.787</td>
</tr>
<tr>
<td></td>
<td>R² 0.093</td>
<td>0.107</td>
<td>0.116</td>
</tr>
<tr>
<td></td>
<td>p&gt;χ² 0.167</td>
<td>0.072</td>
<td>0.262</td>
</tr>
<tr>
<td>Discontinuity</td>
<td>1.920</td>
<td>-1.160</td>
<td>2.620</td>
</tr>
<tr>
<td></td>
<td>p 0.562</td>
<td>0.624</td>
<td>0.348</td>
</tr>
<tr>
<td></td>
<td>bw 57.630</td>
<td>47.290</td>
<td>42.480</td>
</tr>
<tr>
<td>N</td>
<td>200/202</td>
<td>201/203</td>
<td>144/145</td>
</tr>
<tr>
<td>Obfuscation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linear Estimate</td>
<td>0.040</td>
<td>-0.017</td>
<td>-0.075</td>
</tr>
<tr>
<td></td>
<td>p 0.872</td>
<td>0.934</td>
<td>0.614</td>
</tr>
<tr>
<td></td>
<td>R² 0.068</td>
<td>0.109</td>
<td>0.095</td>
</tr>
<tr>
<td></td>
<td>p&gt;χ² 0.486</td>
<td>0.063</td>
<td>0.489</td>
</tr>
<tr>
<td>Discontinuity</td>
<td>0.570</td>
<td>0.429</td>
<td>1.070</td>
</tr>
<tr>
<td></td>
<td>p 0.271</td>
<td>0.388</td>
<td>0.192</td>
</tr>
<tr>
<td></td>
<td>bw 82.760</td>
<td>90.630</td>
<td>33.070</td>
</tr>
<tr>
<td>N</td>
<td>200/202</td>
<td>201/203</td>
<td>144/145</td>
</tr>
</tbody>
</table>

Notes: The linear estimates are the agency-level random effects GLS estimates for the variable prior/subsequent, an indicator for whether the observation occurred in the six months before or six months after each decision. Shaded cells indicate that the direction of the coefficient matches the theoretical prediction in Table 1. Reported below the linear estimates are the p-value, the R² (a measure of model fit), and the p>χ² (the probability that the model as a whole is significant). The discontinuity treatment effect, τ, uses a local-linear regression model, a triangular kernel function, and data-driven bandwidth selection of on average about two months. See Sebastian Calonico, Matias D. Cattaneo, & Rocío Titiunik, Robust Data-Driven Inference in the Regression Discontinuity Design, 14 Stat. J. 909 (2014). The discontinuity estimate is the change in the local-linear estimate for the bandwidth (bw) indicated. Statistically significant results are denoted as follows: * p<.1, ** p<.05, *** p<.01, **** p<.001. Each model included a vector of covariates measuring characteristics of the rule, including original, ruleage, remand, priority, major, regplan, smallbiz, and federalism.
Figure 3: Effect of Auer-Strengthening Decisions on Rule Text

Figure 4: Effect of Auer-Strengthening Decisions on Preambles

Notes: Regression discontinuity results on either side of the time of decision are obtained by local-linear regression using a triangular kernel function and with a bandwidth of 200 days. Confidence intervals are at 95 percent with open squares indicating the local average of the observations within each bin.
Table 3, along with Figures 5 & 6, repeat the same analysis, but only with respect to the cases in the study that were theorized to have signaled a diminution in the vitality of the Auer-deference regime: i.e., to have made it more likely that courts would subject any interpretive claim by an agency to greater scrutiny, perhaps by carving out another exception to Auer deference or by weakening it or overturning it in the next case. These cases include Gonzales, Talk America, and Christopher, all of which either carved out exceptions to Auer or signaled a growing discontent on the Court with the status quo.

Again, the models fail to yield any strong evidence consistent with the perverse incentives theory. While there are a few isolated exceptions, the estimates are again primarily statistically insignificant and frequently at odds with theoretical expectations. In those few instances where the estimates are statistically significant—e.g., on both estimates for the Flesch-Kincaid Grade Level and the discontinuity estimate for cognitive complexity for preambles associated with proposed rules—they actually suggest the opposite direction of change to what the perverse incentives theory would predict. In each instance, the evidence says—with a fair degree of confidence—that the agency wrote with greater opacity after receiving a signal from the court that Auer deference was on shaky ground. It is difficult to square these kinds of results with the perverse incentives theory, which boldly claims that agencies take advantage of Auer deference. If agencies write with greater opacity even when Auer’s vitality is at issue, then even if real, the perverse incentives in practice cannot have enough bite to be of any practical significance.
Table 3: Estimates of the Effect of Auer-Weakening Decisions on Proxies for Ambiguity

<table>
<thead>
<tr>
<th>Word Count</th>
<th>Final Rules</th>
<th>Proposed Rules</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Rule Text</td>
<td>Preamble</td>
<td>Rule Text</td>
</tr>
<tr>
<td><strong>Linear Estimate</strong></td>
<td>-0.005</td>
<td>0.000</td>
<td>-0.003</td>
</tr>
<tr>
<td>p</td>
<td>0.501</td>
<td>0.820</td>
<td>0.689</td>
</tr>
<tr>
<td>R²</td>
<td>0.247</td>
<td>0.335</td>
<td>0.152</td>
</tr>
<tr>
<td>p &gt; χ²</td>
<td>0.000</td>
<td>0.000</td>
<td>0.135</td>
</tr>
<tr>
<td>Discontinuity</td>
<td>0.033</td>
<td>0.010*</td>
<td>-0.007</td>
</tr>
<tr>
<td>p</td>
<td>0.159</td>
<td>0.030</td>
<td>0.792</td>
</tr>
<tr>
<td>bw</td>
<td>61.460</td>
<td>51.554</td>
<td>55.088</td>
</tr>
<tr>
<td>N</td>
<td>143/149</td>
<td>143/149</td>
<td>126/133</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>F-K Grade Level</th>
<th>Final Rules</th>
<th>Proposed Rules</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Rule Text</td>
<td>Preamble</td>
<td>Rule Text</td>
</tr>
<tr>
<td><strong>Linear Estimate</strong></td>
<td>0.127</td>
<td>0.099</td>
<td>0.514</td>
</tr>
<tr>
<td>p</td>
<td>0.829</td>
<td>0.675</td>
<td>0.316</td>
</tr>
<tr>
<td>R²</td>
<td>0.247</td>
<td>0.183</td>
<td>0.103</td>
</tr>
<tr>
<td>p &gt; χ²</td>
<td>0.000</td>
<td>0.002</td>
<td>0.549</td>
</tr>
<tr>
<td><strong>Discontinuity</strong></td>
<td>-3.690</td>
<td>-0.192</td>
<td>0.819</td>
</tr>
<tr>
<td>p</td>
<td>0.139</td>
<td>0.870</td>
<td>0.749</td>
</tr>
<tr>
<td>bw</td>
<td>40.760</td>
<td>60.369</td>
<td>60.398</td>
</tr>
<tr>
<td>N</td>
<td>143/149</td>
<td>143/149</td>
<td>126/133</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cognitive Complexity</th>
<th>Final Rules</th>
<th>Proposed Rules</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Rule Text</td>
<td>Preamble</td>
<td>Rule Text</td>
</tr>
<tr>
<td><strong>Linear Estimate</strong></td>
<td>-0.968</td>
<td>0.099</td>
<td>-0.101</td>
</tr>
<tr>
<td>p</td>
<td>0.183</td>
<td>0.879</td>
<td>0.881</td>
</tr>
<tr>
<td>R²</td>
<td>0.234</td>
<td>0.079</td>
<td>0.194</td>
</tr>
<tr>
<td>p &gt; χ²</td>
<td>0.000</td>
<td>0.692</td>
<td>0.021</td>
</tr>
<tr>
<td><strong>Discontinuity</strong></td>
<td>0.267</td>
<td>-2.386</td>
<td>2.828</td>
</tr>
<tr>
<td>p</td>
<td>0.946</td>
<td>0.335</td>
<td>0.194</td>
</tr>
<tr>
<td>bw</td>
<td>55.403</td>
<td>48.891</td>
<td>55.501</td>
</tr>
<tr>
<td>N</td>
<td>143/149</td>
<td>143/149</td>
<td>126/133</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Obfuscation</th>
<th>Final Rules</th>
<th>Proposed Rules</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Rule Text</td>
<td>Preamble</td>
<td>Rule Text</td>
</tr>
<tr>
<td><strong>Linear Estimate</strong></td>
<td>-0.063</td>
<td>-0.324</td>
<td>-0.206</td>
</tr>
<tr>
<td>p</td>
<td>0.825</td>
<td>0.249</td>
<td>0.352</td>
</tr>
<tr>
<td>R²</td>
<td>0.115</td>
<td>0.099</td>
<td>0.151</td>
</tr>
<tr>
<td>p &gt; χ²</td>
<td>0.273</td>
<td>0.443</td>
<td>0.139</td>
</tr>
<tr>
<td><strong>Discontinuity</strong></td>
<td>0.819</td>
<td>-0.337</td>
<td>-0.572</td>
</tr>
<tr>
<td>p</td>
<td>0.413</td>
<td>0.722</td>
<td>0.547</td>
</tr>
<tr>
<td>bw</td>
<td>35.891</td>
<td>42.026</td>
<td>54.720</td>
</tr>
<tr>
<td>N</td>
<td>143/149</td>
<td>143/149</td>
<td>126/133</td>
</tr>
</tbody>
</table>

Notes: The linear estimates are the agency-level random effects GLS estimates for the variable prior/subsequent, an indicator for whether the observation occurred in the six months before or six months after each decision. Shaded cells indicate that the direction of the coefficient matches the theoretical prediction in Table 1. Reported below the linear estimates are the p-value, the R² (a measure of model fit), and the p > χ² (the probability that the model as a whole is significant). The discontinuity treatment effect, τ, uses a local-linear regression model, a triangular kernel function, and data-driven bandwidth selection of on average about two months. See Sebastian Calonico, Matias D. Cattaneo, & Rocío Titiunik, Robust Data-Driven Inference in the Regression Discontinuity Design, 14 State J. 909 (2014). The discontinuity estimate is the change in the local linear estimate for the bandwidth (bw) indicated. Statistically significant results are denoted as follows: ^ = p < .1, * = p < .05, ** = p < .01, *** = p < .001. Each model included a vector of covariates measuring characteristics of the rule, including original, ruleage, remand, priority, major, regplan, smallbiz, and federalism.
Figure 5: Effect of Auer-Weakening Decisions on Rule Text

Figure 6: Effect of Auer-Weakening Decisions on Preambles

Notes: Regression discontinuity results on either side of the time of decision are obtained by local-linear regression using a triangular kernel function and with a bandwidth of 200 days. Confidence intervals are at 95 percent with open squares indicating the local average of the observations within each bin.
In short, where the predicted change plays out, the *Auer* effect is both substantively and statistically insignificant, and the instances of convergence between theory and reality are not consistent enough to draw any inferences in favor of the hypothesis. That is, while we cannot say that the effect is truly zero,\textsuperscript{165} we can definitely say that the evidence here does not bear out the worst fears about widespread exploitation of *Auer*’s alleged perverse incentives, at least with respect to the aggregated data. We now turn to a more fine-grained analysis broken out by cases, where the results are similarly confounding.

B. Individual Analysis

As previously mentioned, the aggregated data may obscure an effect that is confined to a handful of decisions. Perhaps these Supreme Court decisions are not all alike. Perhaps the signal is considerably stronger in the aftermath of one decision than it is in another. These are reasonable concerns, especially given empirical studies that show that only a fraction of agency rule writers have even heard of *Auer* deference\textsuperscript{166} and given that the concerted critique of *Auer* (and the real threat of doing away with it) only picked up steam in recent years. In other words, cases like *Auer*, *Gonzales*, and *Coeur Alaska*, which involved more or less routine applications of the *Auer* framework and little discussion of the overall fate of the deference regime, may not have been interpreted or even recognized by a critical mass of agency rule writers as important to the agency’s work. With cases like *Talk America*, *Christopher*, *Perez*, and *United Student Aid Fund*, in contrast, the vitality of the *Auer* regime was a topic of much discussion. For these reasons, it is important to repeat the analysis at the individual case level.

Table 4 presents the estimates of the regression discontinuity treatment effect for each of the proxies for ambiguity, broken out by the Supreme Court decision at the center of each window and by the different components of the notice (i.e., the rule text to be included in the Code of Federal Regulations and the preamble text that explains the agency’s action). Overall, measured by the baseline of the theoretical expectations outlined in Table 1, the results do not provide strong evidence of a perverse incentives effect. As with the prior results in Tables 2 & 3, the signage is almost as good as random (compared with the case-level predictions in Table 1), and few of the results approach statistical significance.


\textsuperscript{166} See Walker, *supra* note 157, at 1007.
| Table 4: Estimates of the Effect of Auer-Related Decisions on Agency Rule Writing |
|---------------------------------|----------------|----------------|----------------|
|                                 | Word Count    | F-K            | Cog Complex    | Obfuscation    |
| Auer (n= 89)                    |               |                |                |                |
| Rule                            | -.007 (.606)  | -.210* (.024)  | .552 (.692)    | .780^ (.096)   |
| Preamble                        | .005 (.307)   | -.858 (.242)   | .276 (.852)    | .397 (.194)    |
| Gonzales (n=68)                 |               |                |                |                |
| Rule                            | .010 (.557)   | 1.49 (.294)    | 1.35 (.405)    | .008 (.986)    |
| Preamble                        | .006* (.029)  | .476 (.226)    | .547 (.730)    | .055 (.899)    |
| Coeur Alaska (n=92)             |               |                |                |                |
| Rule                            | -.007 (.544)  | .952 (.489)    | 2.24 (.327)    | -.278 (.682)   |
| Preamble                        | .008 (.192)   | -.711 (.167)   | 2.08 (.280)    | .653 (.146)    |
| Talk America (n=133)            |               |                |                |                |
| Rule                            | -.010 (.620)  | .587 (.579)    | .596 (.756)    | -.152 (.731)   |
| Preamble                        | .005 (.169)   | -.254 (.690)   | 2.05 (.132)    | -.371 (.256)   |
| Christopher (n=81)              |               |                |                |                |
| Rule                            | -.023 (.218)  | 1.77 (.481)    | -3.61 (.160)   | -.916* (.042)  |
| Preamble                        | -.003 (.470)  | -.075 (.919)   | -3.23* (.027)  | -.112** (.010) |
| Perez (n=97)                    |               |                |                |                |
| Rule                            | .007 (.335)   | .068 (.931)    | .123 (.993)    | .107 (.789)    |
| Preamble                        | .000 (.968)   | .860 (.179)    | .863 (.574)    | .242 (.647)    |
| United Student (n= 70)          |               |                |                |                |
| Rule                            | -.008 (.341)  | .749 (.395)    | 1.64 (.438)    | .293 (.501)    |
| Preamble                        | .001 (.869)   | -.193 (.803)   | 1.70 (.450)    | -.041 (.964)   |

Notes: The table presents estimates of the discontinuity, τ, between covariate-adjusted local-linear regressions using a standard bandwidth of 180 days before and after the relevant decision. Estimates are acquired using a triangular kernel function. The included covariates are type, original, remand, priority, major, regplan, smallbiz, and federalism. Statistically significant results are denoted as follows: ^ = p<.1, * = p<.05, ** = p<.01, *** = p<.001.

There is, however, one exception. The results on Christopher v. SmithKline Beecham for the cognitive complexity and obfuscation variables are correctly signed and statistically significant. Around that case, which greatly expanded the probability that an agency interpretation would be stringently reviewed where the interpretation created “unfair surprise” for regulated entities, the agencies in the study shifted fairly dramatically to a less opaque writing style both in the rule text and in the associated preambles, at least on two of the measures of opacity. These jumps can be seen clearly in Figure 7, which plots the treatment effect for the period after Christopher for these two proxies for ambiguity. One should nevertheless be cautious in interpreting these results as evidence of the operation of the perverse incentives theory. After all, several of the estimates of the effect of the decision on other measures are incorrectly signed, showing decreasing clarity in the aftermath of Christopher.

---

However, focusing just on the last two proxies for ambiguity—cognitive complexity and obfuscation—there is something of a longitudinal story to be told in Table 4. Indeed, while these data are not suited to provide any useful time-series analysis over the course of decades, it is notable that the estimates start positive in the early cases when the Auer regime was only beginning to be questioned in cases like Gonzales, turn negative and then sharply negative at the apogee of Auer skepticism at the Supreme Court in cases like Talk America and Christopher, and then arrive at a denouement with a return to mixed and mostly positive (but, to be sure, statistically insignificant) results in cases like Perez and United Student Aid Fund, where the Court was widely perceived to have stepped back from the breach, at least temporarily. This story of ebbs and flows looks facially plausible, and it accords with some of our intuitions about the overall arc of change in the Auer arena. But, critically, it is nearly impossible to disentangle the confounding factors that might also account for changes over the course of this 20-year period. It is far better to stick with the regression discontinuity approach and view each case—or like cases—as a snapshot, and in that case there is little consistency or strength in the evidence of any perverse incentives effect.

168 See Barmore, supra note 11, at 821 (noting that “the first serious sign” that the Court was interested in overturning Auer came in Talk America in 2011).
IV.  DISCUSSION

The foregoing empirical analysis provides a number of insights—some doctrinal, some theoretical, and some methodological—for current debates in administrative law. In this Part, I review some of these critical takeaways. Beyond the simple point that agencies do not appear to act on any perverse incentives created by the Auer rule (a finding which considerably simplifies the debate over Auer), the study has a number of implications for administrative law. For instance, the lack of any evidence of insincerity or opportunism in regulatory draftsmanship bears on the growing debate about the proper methodology of regulatory interpretation. Moreover, the same lack of evidence of insincerity or opportunism serves as yet another potent reminder of the limitations of incentive-based arguments in administrative law. I address each point in turn.

A. The Future of Auer Deference

As should be clear, the empirical analysis casts serious doubt on a prominent critique of Auer deference—that agencies promulgate ambiguous legislative rules in order to maximize their interpretive discretion as they go about informally applying the rules they have written. Put simply, there is no systematic evidence that agencies do this. The doctrinal significance of this finding is plain to see. Insofar as the strength of the critique of Auer rests on a misappraisal of the actual effect of the allegedly perverse incentive structure, there is little reason for the Supreme Court to consider doing away with the doctrine.169 It may still consider ways to define the scope of Auer deference, including by fleshing out the interpretive methodology that triggers Auer deference in the first place.170 But doing away with Auer deference entirely would be a gross overreaction.

At the same time, the lack of any evidence that agencies respond to changes to the strength of the Auer doctrine suggests that Auer’s defenders would have little to worry about if Auer were significantly scaled back or even replaced with something like Skidmore deference. The precise level of deference agencies receive in their interpretations of regulatory text appears to be somewhat epiphenomenal to the considerations that drive the timing of interpretive specification. Regardless of the level of deference they receive in their interpretive rules, agencies will likely continue to promulgate rules that capture the level of precision that the agency feels is optimal, given the circumstances.171

A quick word is in order about what the evidence does not do or say. Because the focus is on rule writing and not on agencies’ actual invocation of interpretive authority in guidances and the like, the study has essentially nothing to say about the distinct argument that Auer deference encourages agencies to rely on informal interpretations rather than notice-and-comment rulemaking. In this variation of the critique, the denominator—i.e., the clarity of the rules—is assumed to be exogenously set. However much ambiguity there is in the rules as written

169 This, of course, assumes that the arguments about the suboptimal allocation of interpretive authority between courts and agencies discussed supra notes 25-27 and accompanying text. But that seems like quite a weak argument. See supra Part I.A.
170 See infra Part IV.B.
171 Diver, supra note 36.
much genuine discretion the agency possesses. Assuming that the agency can identify some ambiguity, *Auer* changes the numerator, or the expected payoff for invoking interpretive authority to resolve the ambiguity. As the argument goes, however much ambiguity may exist in the rules, the more or less flat rule of deference in *Auer* (as opposed to the more contextualized considerations that might drive a court to afford *Skidmore* weight to the interpretation) will encourage agencies to take the gamble more often than they otherwise would, leading to a general erosion of administrative law’s long-standing commitment to legislative rulemaking. This separate argument is an important one that deserves its own close scrutiny. Indeed, several empirical explorations have begun to examine whether it holds: i.e., whether agencies strategically resort to guidance to insulate themselves from procedural controls. The answer, as of yet, appears to be no, but more work remains to be done.

The important thing to note here, however, is that the opportunity for arbitrage is separate, and, as a theoretical matter, perhaps far more concerning. Sticking with the fractional imagery, what is at issue with the perverse incentives argument is an agency’s unilateral control over the denominator, or the ability to exponentially increase the opportunities to regulate by guidance. Yet, as we have seen, even under the best conditions to find such behavior, the evidence does not bear out the worst fears.

Why this is the case is more difficult question to answer, and this article does not venture too far into the unknown. Perhaps it has something to do with cross-cutting agency incentives that overwhelm the impulse; or perhaps with ignorance of *Auer* deference among rule writers. It may also be constrained by the fact that promulgating “mush” simply would not fly with participants in notice-and-comment rulemaking, who are as primed to challenge vague rules as arbitrary or unconstitutional as they are specific rules. For a regulated entity that wants to plan its compliance, just as for a public interest group that wants to ensure effective enforcement of a rule, there will often be enormous pressure on an agency to be more specific. Indeed, much of the “back-and-forth with external commenters and political monitors” in the preambles accompanying legislative rules focuses on “respond[ing] to public comments raising potential ambiguities in the proposed regulation.”

Whatever the reason for the results, the key takeaway from this study is that the particular brand of perverse incentives argument that Supreme Court has relied on heavily to critique *Auer* is unsubstantiated.

---

173 On one account, it is far more concerning because the arbitrage amounts to agencies delegating themselves more power than they otherwise have. Stephenson & Pogoriler, *supra* note 4, at 1471. On the other, as Adrian Vermeule points out, the agency is still bound by the four corners of the statutory delegation—even an ambiguous phrase in a previously promulgated regulation might be challenged as inconsistent with the governing statute. Vermeule, *Law’s Abnegation*, *supra* note 2.
174 Paralyzed Veterans of Am. V. D.C. Arena L.P., 117 F.3d 579, 584 (D.C. Cir. 1997).
175 Nou, *supra* note 3, at 85.
B. Methods of Regulatory Interpretation

One of the more fruitful developments coming from the debate over Auer is a growing attention to the unique challenges posed when courts interpret regulations. While the literature on statutory interpretation is vast and a seeming consensus has emerged over some primary commitment to a textualist philosophy and methodology,\(^\text{176}\) it is not at all clear that the same kind of considerations should prevail when it comes to determining the meaning of a regulation. Regulations are undoubtedly law,\(^\text{177}\) and, under a certain theory, simply an extension of the Article II power to promulgate statutes, but they are distinctive enough in their purpose and operation that a growing number of scholars have considered a stand-alone methodology for regulatory interpretation.\(^\text{178}\) Several of these accounts seek to rehabilitate methodologies long since out of vogue in the statutory context. Kevin Stack, for instance, pushes for a “purposivist” theory of regulation interpretation based on the fit between the institutional realities of regulatory policymaking and the assumptions of the Legal Process School.\(^\text{179}\) Even intentionalism, which long ago succumbed to the critique of social choice theorists that the underlying concept of institutions having an identifiable “intent” is incoherent,\(^\text{180}\) has defenders in the context of regulatory interpretation.\(^\text{181}\) The general thrust of these dynamic methodologies is that the

---

\(^{176}\) William Baude & Ryan Doerfler, The (Not So) Plain Meaning Rule, 84 U. Chi. L. Rev. at 1 (forthcoming 2017) (“Many tenets of statutory interpretation take a peculiar form. They allow consideration of outside information—legislative history, practical consequences, the statute’s title, etc.—but only if the statute’s text is unclear or ambiguous.”); William N. Eskridge, Jr., The New Textualism, 37 UCLA L. Rev. 621, 660 (1990). More generally, a group of “accommodationists” argue that a core of common concerns and techniques underlie all of the major approaches to statutory interpretation, including textualism, intentionalism, and purposivism. See Jonathan R. Siegel, The Inexorable Radicalization of Textualism, 158 U. Pa. L. Rev. 117, 119 (2009) (“The latest move in the interpretation wars, however, is to declare something of a truce.”); see also id. at 120 (noting that at least one accommodationist views “rival methods” as having “the same underlying goal”) (emphasis in original).

\(^{177}\) Stack, Interpreting Regulations, supra note 22, at 357 (“While all agree that regulations are primary sources of law, strikingly little attention has been devoted to the method of their interpretation.”); see also Stack, Preambles as Guidance, supra note 4.

\(^{178}\) Stack, Interpreting Regulations, supra note 22, at 358 (“[A] theory of regulatory interpretation must be grounded in the distinctive character of regulations and the institutions that issue them.”).

\(^{179}\) Stack, Interpreting Regulations, supra note 22, at 363 (discussing Henry M. Hart, Jr. & Albert M. Sacks, The Legal Process (William N. Eskridge, Jr. & Philip P. Frickey eds. 1994)). In essence, this approach calls on judges to interpret the meaning of regulatory text in light of an assumption that regulators are “reasonable persons, pursuing reasonable purposes within the permissible range of their discretion,” and to use preambles as a guide to exploring these reasonable purposes.” See id. at 397-98. Others go even further, assuming that virtually any document in the rulemaking record could support the purposivist inquiry. See Russell L. Weaver, Judicial Interpretation of Administrative Regulations: An Overview, 53 U. Cinn. L. Rev. 681, 696 (1984).


preambles that accompany rules and explain the agency’s thinking ought to be considered as a source in determining the meaning of a regulation.

In contrast, in a recent article, Jennifer Nou seeks to bring the methodology of regulatory interpretation more into alignment with the textualist approach that is ascendant in the field of statutory interpretation. Recognizing that the appropriate methodology turns on questions of institutional capacity and competence, Nou argues that purposivism and intentionalism are ultimately poorly calibrated to optimize judicial review, and that an enlightened form of textualism would strike a better balance. In large part, this critique is based on the notion that the “public meaning” of a regulation is determined by an agreement among pivotal actors. If ascertaining this public meaning is the goal of regulatory interpretation, then it follows inexorably that “judges should rely on materials that are likely to be sincere, as opposed to strategic attempts to misstate the terms of the agreement.”

It is perhaps no surprise, then, that Nou spends much of her article outlining what she calls the “strategic self-delegation” problem and teasing out its implications for the choice of methodology. This problem, which is simply an extension of the perverse-incentives logic in the Auer debate, counsels an interpretive methodology that “can reduce the discretion available to subsequent interpreters” so that it can “reduce an agency’s ex ante ability to issue overly broad regulatory texts.” According to Nou, “[t]he concern is particularly acute for those interpretive approaches that invite broad consultation of the regulatory history, as the modern administrative record often spans different and likely conflicting administrators and various levels of an agency’s hierarchy,” but it also exists where, as in Stack’s purposivism, a judge makes “synthetic” judgments about purpose evinced by the entire preamble. Given these priors, Nou offers a significantly curtailed methodology, sounding largely in textualism, that rank orders interpretative materials based on how likely they are to be credible, as assured by having undergone publication and “multiple forms of oversight by pivotal actors.” For Nou, this means that “provision-by-provision analyses in preambles are the most reliable sources of the text’s public meaning” and any accompanying regulatory analysis a second-best source, but that the more general statements of basis and purpose in preambles are “pitched at too high a level of abstraction to inform the court’s specific interpretive task.”

182 Nou, supra note 3, at 94.
183 Id. at 86; id. at 117 (“Positive political theories developed in the statutory context invite a conception of regulations as bargains struck between various regulatory actors and interest groups at different veto-gates of the drafting process—involving the agency head, the President during OIRA-coordinated review, and legislators pursuant to the CRA.”).
184 Id. at 86.
185 Id. at 104.
186 Id. at 104.
187 Id. at 105.
188 Id. at 118.
189 Id. at 119
190 Id. at 126.
191 Id. at 120.
Nou’s attempt to cabin regulatory interpretation and tie the methodology mostly to the text of rules is in many ways an effort to launder a response to the perverse incentives critique of Auer deference through Auer’s emerging first step. By incorporating a concern for the danger of insincerity into the very methodology for determining when an agency’s interpretation is “plainly erroneous” or “inconsistent” with the regulatory text, Nou is able to save Auer deference, albeit with a twist. What is missing—and what the empirical analysis in this study enables—is recognition of the fact that the concerns about insincerity are probably misplaced even at this first stage of the debate. Concerns about self-delegation run right through this debate over the methodology of regulatory interpretation, usually resulting in some reticence about courts’ reliance on Federal Register preambles to interpret regulatory text. But there is no evidence that the preambles agencies are drafting are Trojan horses. Although Table 5 demonstrates that Auer-related decisions do sometimes have a visible effect on the extent to which the agencies use capacious preambles to provide texture to their rules, the direction of the changes vis-à-vis changes to the doctrine again complicates the perverse incentives theory. The results in Table 5 seem to suggest that Auer-strengthening decisions consistently lead to a reduction in the opacity of preambles—a result completely at odds with the traditional form of the perverse incentives theory. Moreover, if anything, Auer-weakening decisions, which should result in a decrease in preamble opacity relative to rule texts, appear to cause an increase in relative opacity.

Table 5: Effect of Auer-Related Decisions on Relative Preamble Ambiguity

<table>
<thead>
<tr>
<th>(Preamble-Rule)</th>
<th>60-Day RD Estimate</th>
<th>180-Day RD Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All Auer-</td>
<td>All Auer-</td>
</tr>
<tr>
<td></td>
<td>Strengthening</td>
<td>Weakening</td>
</tr>
<tr>
<td>Word Count</td>
<td>-.012 (.184)</td>
<td>0.014 (.497)</td>
</tr>
<tr>
<td>FK Grade Level</td>
<td>-.723 (.509)</td>
<td>0.552 (.766)</td>
</tr>
<tr>
<td>Cog Complexity</td>
<td>3.29^ (.068)</td>
<td>-.757 (.834)</td>
</tr>
<tr>
<td>Obsfuscation</td>
<td>.201 (.708)</td>
<td>0.547 (.468)</td>
</tr>
<tr>
<td>ABS (Preamble-Rule)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Word Count</td>
<td>-.010 (.265)</td>
<td>.021 (.232)</td>
</tr>
<tr>
<td>FK Grade Level</td>
<td>-1.08 (.201)</td>
<td>.463 (.802)</td>
</tr>
<tr>
<td>Cog Complexity</td>
<td>-.819 (.431)</td>
<td>3.93** (.003)</td>
</tr>
<tr>
<td>Obsfuscation</td>
<td>-.155 (.720)</td>
<td>.511 (.273)</td>
</tr>
</tbody>
</table>

Notes: The table shows the estimates of the discontinuity, τ, between local linear regressions using a standard bandwidth of 180 days before and after the relevant decision for two sets of dependent variables: 1) the simple difference between scores for each proxy for ambiguity for preambles and the associated rules, and 2) the absolute value of that difference. Estimates are acquired using a triangular kernel function. The included covariates are type, original, remand, priority, major, regplan, smallbiz, and federalism. Statistically significant results are denoted as follows: ^ = p<.1, * = p<.05, ** = p<.01, *** = p<.001.

192 Cf. id. at 142-43.
193 See id. at 148 (outlining an “intermediate approach between Seminole Rock and Skidmore deference” that “create[s] strong incentives for clear regulatory drafting and suppl[i]es a stronger judicial check on a subsequent agency’s interpretation of the text’s public meaning”).
To be sure, many of Nou’s concerns about a dynamic form of regulatory interpretation still exist. Any such approach is still largely based on an almost indefensible legal fiction: the identification of a single regulatory “intent” or “purpose in what is decidedly a messy process of rulemaking.”\textsuperscript{194} And, more fundamentally, preambles are undoubtedly written more capacious than rules,\textsuperscript{195} and that probably does give agencies a great deal of room to maneuver in subsequent interpretations.\textsuperscript{196} The larger point to underscore here, however, is that, just as in debates over the fate of Auer deference, much of the strength of the argument for a constrained interpretive methodology in the regulatory context comes from big theoretical claims about perverse incentives and self-delegation. And, just as in the context of debates over the fate of Auer deference, we ought to proceed very cautiously until the evidence converges with our theoretical models. So far, the evidence is mixed and weak at best.

C. The Limits of Incentives-Based Arguments

Administrative law scholarship is fond of the concept of incentives, and it frequently asks them to do more work than they can actually do. A survey of the literature reveals pervasive arguments that this or that institutional change will realign various constituencies’ incentives.\textsuperscript{197} These arguments then tend to result in protracted discussions of the policy tradeoffs of various institutional arrangements, and sometimes the arguments become so self-referential that the underlying assumption of a relationship between a set of incentives and actual behavior becomes reified. This, arguably, is the case with the debate over Auer. The perverse incentives hypothesis has become the perverse incentives orthodoxy, with major implications for administrative law.

\textsuperscript{194} See Daniel A. Farber & Anne Joseph O’Connell, The Lost World of Administrative Law, 92 Tex. L. Rev. 1137 (2014) ("Our thesis is simple but powerful: the actual workings of the administrative state have increasingly diverged from the assumptions animating the APA and classic judicial decisions that followed.").

\textsuperscript{195} See infra Methodological Appendix.

\textsuperscript{196} See Nou, supra note 3, at 120 (“By contrast to Stack’s purposivist approach, regulatory textualism rejects reliance on the broad statements of purpose often found in preambles in favor of the more specific explanatory provisions. Such broad statements are pitched at too high a level of abstraction to inform the court’s specific interpretive task.”). But see Stack, Preambles as Guidance, supra note 4, at 1272-77 (discussing the procedural constraints on preamble drafting and arguing that preambles “have an authoritative legal status that separately-issued guidance does not”).

\textsuperscript{197} Cary Coglianese, Empirical Analysis and Administrative Law, 2002 U. Ill. L. Rev. 1112, 1113 (“Administrative law is constructed and reconstructed on the basis of assumptions about how particular procedural arrangements will affect the behavior and performance of government officials and organizations.”). Relevant examples of incentives-based arguments include, just to name a few, assertions that parts of the Freedom of Information Act (FOIA) encourage agencies to go off the record and classify more information than they otherwise would, see David Pozen, Freedom of Information Beyond the Freedom of Information Act, at 23 (Feb. 27, 2017), available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2907719; Adam Candeub, Transparency in the Administrative State, 51 Hous. L. Rev. 385, 393 (2013), and the theory that certain judicial doctrines undermine positive incentives created by overlapping and underlapping jurisdictional delegations in statutes, see Gersen, supra note 158, at 3.
Whether the empirical analysis above effectively debunks the perverse incentives hypothesis is less important than that it destabilizes the perverse incentives orthodoxy. As I have explained, I too find the perverse incentives logic intuitively plausible, but plausibility is insufficient, especially when reform agendas are built on it and nothing more. Rules and institutions do have consequences for behavior, and making fundamental changes to them without understanding empirically what is going on is to beg for a slew of unintended consequences that we cannot even begin to understand.

As it turns out, we have seen this story before. For example, for decades it was received wisdom in the field of administrative law that judicial review of agency rules stifles the rulemaking process. This “ossification” thesis claimed that agencies respond to the threat of vacatur by avoiding rulemaking altogether.198 In other words, this influential perspective posited the existence of a certain set of incentives, built on the back of a certain caricature of regulators as prone to shirking their duties, that resulted in systematically changed behavior at odds with the public interest.199 The ossification thesis was not merely an academic exercise; it undergirded a reform agenda that sought to free agencies from the grip of the perverse incentives structure by scaling back the stringency of hard-look review,200 relieving some of the procedural burdens of rulemaking,201 and releasing some of the adversarial pressure that might lead to litigation.202 Despite the near consensus, the evidence of ossification was almost always anecdotal.203

More recently, systematic empirical analyses have revealed that “evidence that ossification is either a serious or widespread problem is mixed and relatively weak,” and that, even at the height of the “allegedly ossified era,” “federal agencies remain[ed] able to propose and promulgate

---


201 McGarity, supra note 199.


203 See Coglianese, supra note 197, at 1127 (“[A]dministrative law scholars have failed generally to produce systematic empirical analysis of the effects of judicial review.”); Mashaw & Harfst, supra note __, at 275 (“[T]he normative expectations of administrative lawyers have seldom been subjected to empirical verification of a more than anecdotal sort.”).
historically large numbers of regulations, and to do so relatively quickly.” To be sure, even this systematic evidence does not purport to rule out the possibility that particular regulatory initiatives succumbed to procedural pressure. There will always be such anecdotes, and it may even be the case that the ossification thesis is correct, despite the best currently available evidence. The takeaway point is that systematic scrutiny has significantly undermined our confidence in what was once almost a tautology—that judicial review deters rulemaking.

It would be haughty and cursory to suggest that all incentives-based arguments about the administrative process are unlikely to be true, but it is also undoubtedly true that the process of regulatory policymaking is extraordinarily complex. The ossification thesis focused inordinately on a lawyer’s conception of risk, and in the process ignored the many incentives that agencies have to continue doing business as usual regardless of what agency lawyers say about the probability of vacatur. Similarly, if the evidence in this empirical study of Auer’s incentives is to be believed (or, more precisely, preliminarily accepted as evidence in favor of rejecting the null hypothesis that Auer has no effect on regulatory draftsmanship), the mistaken confidence can be understood as the result of an unrealistic lawyer-centrism—one that gives inordinate attention to strategy and too little attention to the realities of regulatory policymaking. In short, we ought to heed these warnings and be more wary about casually accepting empirically untested incentives-based arguments and altering doctrine in response.

CONCLUSION

As Justice Thomas noted in his dissent from the denial of a certiorari petition in United Student Aid Funds, Auer deference appears to be “on its last gasp.” Predicting when Auer will meet its demise is a popular pastime in administrative law scholarship, particularly after the fundamental shift in power in Washington, D.C. following the 2016 election. But the diagnosis is premature; not all of the evidence is in. Concerns about Auer rest on a hypothesis about the real-world effect of the perverse incentive structure that the deference regime creates, and we have only begun to test whether our theoretical suppositions match reality on the ground. So far, the evidence does not bear out the hypothesis. More work remains to be done, but, for now, Auer’s vital signs are still measurable for anyone who is checking.

204 Yackee & Yackee, supra note 199, at 1421.
205 Yackee & Yackee, supra note 199, at 1422.
206 See Richard J. Pierce, Jr., Rulemaking is Real: A Response to Testing the Ossification Thesis, 80 Geo. Wash. L. Rev. 1493 (2012) (arguing that the research by the Yackees is not dispositive and identifying some methodological qualms).
208 See Barmore, supra note 11, at 816 (“It is only a matter of time before the Court decides to revisit Auer.”).
METHODOLOGICAL APPENDIX

This appendix provides greater detail on data collection, the measurement of key variables, and the overall study design.

Sample Identification and Data Collection

For each of the cases included in the study and for each of the agencies selected for study, I sampled notices of proposed rulemaking and notices of final rules in the Federal Register for the six months prior to and subsequent to the publication of the Supreme Court’s decision. For example, for Talk America, which was decided on June 9, 2011, I collected all notices from December 9, 2010 through June 8, 2011, as well as from June 9, 2011 through December 9, 2011. For two of the cases—Talk America and Perez—I also collected the second six months before and after the decision. For example, with Talk America I collected notices published December June 9, 2010 through December 8, 2010 and December 10, 2011 through June 9, 2012. The longer window for these two cases allows for a robustness test on the main results from the 12-month window for most of the cases. Using the advanced search feature on the Federal Register website, I limited the results to those documents included in the Unified Agenda and deemed significant under Executive Order 12,866.

After the sample of notices was generated, a research assistant examined each of the notices, culled only those notices that announced a proposed or final rule and provided regulatory text to be added to the Code of Federal Regulations (CFR), and collected and organized the associated preamble and rule texts. The research assistant then cleaned the text files to prepare them for processing. The cleaning process involved deleting charts and equations, deleting page break markers, and leaving only the raw text and the headers.

Control Covariates

Not even the strongest formulation of the perverse incentives theory posits that all or even most of the potential ambiguity in regulatory texts emerges because of Auer. Rules come in many different forms and attempt to achieve many different results, and some of these differences can be expected independently to cause some of the ambiguities that arise in practice. We might suspect, for instance, that the more substantively important a given rule is, the more ambiguous it is likely to be, because the agency will inevitably have to make compromises and tradeoffs that attempt to satisfy the various interests with a stake in the rule. Or perhaps a rule is likely to be more ambiguous where there are shared state-federal responsibilities, on the theory that the division of authority necessarily creates residual ambiguities over jurisdiction at the very least.

\[A\]

A substantial number of the notices returned in the searches were removed either at this point or during the collection of data on rule characteristics. The notices that were excluded included clerical corrections to existing rules, notices of extension of public comment periods, and other actions that did not result in any proposed amendment or addition to the Code of Federal Regulations.
For each of the notices, which were by design included in the *Unified Agenda*, I used the Regulation Identification Number (RIN) to search the *Unified Agenda* website for a number of rule characteristics, focusing on the entry closest in time to the relevant Supreme Court decision. The *Unified Agenda* contains several fields which agencies are required to fill out when creating an entry. I also logged the issuing agency and the type of notice (proposed rule or final rule)—this information was clearly noted in the *Federal Register* notice itself. Figures A.1 & A.2 show the distribution of the *Unified Agenda* variables by type of notice and by case. These figures show that the distribution of covariate values in the data is remarkably consistent.

**Figure A.1: Frequency of Control Covariates by Type of Notice**

![Figure A.1: Frequency of Control Covariates by Type of Notice](image-url)
Next, I read the introduction to the preamble in order to determine whether the proposed or finalized rule was a new substantive addition to policy or primarily a limited revision to an existing rule (original), as well as whether the proposed or finalized rule was issued in response to a remand from a court or was initiated in response to a granted petition for reconsideration of a previously published rule (remand).

Finally, using the Unified Agenda, I identified each of the sections of the CFR cited. I then searched each Part or Section (however specific the Unified Agenda was) in Cornell University Law School’s online CFR database.210 Using this resource, I was able to log the earliest and latest origin of the cited Parts and Sections for most of the rules in the dataset. I measured the number of years between each origin and the relevant Supreme Court decision and averaged this to produce the measure ruleage. Figure 8 presents the distribution of the ruleage variable, and as it shows clearly, the distribution is bimodal, with a high number of zeros (indicating that the rule text created a brand new Part or Section in the CFR) and second peak around 20 years. This measure serves as a rough proxy for “amendedness,” as rules that have been on the book longer can be expected to have been amended more often.

---

Dependent Variables: Proxies for Vagueness

As discussed above in Part II.C, the study relies on several proxies for ambiguity: the inverse of the natural log of the total word count, the Flesch-Kincaid Grade Level, a cognitive complexity index, and an obfuscation index. Each of these variables is operationalized so that higher scores indicate higher opacity in the text being analyzed.

All of these measures are generated automatically by computer content analysis, which means that they are far more reliable and efficiently replicable than any human coding could ever aspire to achieve.\textsuperscript{211} The tradeoff, however, is that these abstract measures may not be valid—i.e., they may not capture precisely what legal scholars mean when they say that a given statutory or regulatory text is ambiguous. Even after we clarify that the concept of interest is the potential for ambiguity (not actual, or situational ambiguity), much depends on whether the chosen measures actually capture that potential for ambiguity and not some other, related, characteristic of the text. The problem is only heightened by the fact that advocates of the perverse incentives theory have not been clear about what they mean by ambiguity, as well as by the fact that the relationships between legal concepts like ambiguity and vagueness are themselves complicated and unclear.

Each of the measures chosen for inclusion in the study is designed to measure different facets of the clarity of a given text, and the literature has extensively tested the validity of several of

\textsuperscript{211} Black et al., \textit{supra} note 115, at 46-47.
these measures for these purposes. I argue that clarity is closely related to the underlying concept of interest—potential ambiguity, or vagueness. When legal scholars articulate the perverse incentives theory, they do not appear to limit their theory to a narrow, formalistic, and situational concept of syntactic or lexical ambiguity. Instead, they appear to be concerned with a more general notion of semantic slippage, and with the resulting danger of a lack of notice when law is incomprehensible or when a new interpretation is likely to “surprise” regulated entities and the public. The notice problem emerges when agencies promulgate conceptual “mush” that can be molded to support a wide variety of future actions. Measures of general clarity are likely to be valid to capture this sense of semantic slippage, even as they would likely be hopeless for tracking syntactic or lexical ambiguity. And it bears noting that the measures of the rule text load onto a single component in a principal components analysis with an Eigenvalue of 2.38, and this component explains 60 percent of the variance across the measures; likewise, the measures of the preamble text load on two components that explain 59 percent of the variance. This suggests that the measures—as different as they are on the margins—are largely tapping into an underlying, unidimensional concept of clarity.

As an additional effort to demonstrate the face validity of the measures, I look at how the measures perform in explaining textual differences in several so-called anti-parroting cases. Recall that in Gonzales v. Oregon, the U.S. Supreme Court declined to afford Auer deference to the Attorney General’s interpretations of his own regulation, because that regulation did “little more than restate the terms of the statute itself.” As the Court put it, Auer deference is based on the idea that “underlying regulations give specificity to a statutory scheme.” If there is a “near equivalence” between the regulation and the statute, the question “is not the meaning of the regulation but the meaning of the statute,” and the rationale for Auer falls away. The anti-parroting cases can be leveraged to test the validity of my measures of opacity. In theory, if a court declines to afford Auer deference because of the anti-parroting principle, it is saying that the regulation does not add enough specificity or clarity to the statutory language to justify deference, or perhaps even makes the framework more opaque than it was under the statute. If this is true, we can measure both the statute and the regulation in these cases and compare the differences.

Table A.1 conducts this analysis for three cases where the court declined to enforce the anti-parroting principle and three where it did not, including Gonzales itself. The row to focus on is the “Average Δ” row, which computes the mean difference between the statute and

---

212 See supra Part II.C.
213 Barmore, supra note 11, at 818 (“Critics worry that the incentive to promulgate vague regulations would lead to predictably more ambiguous regulations, thereby giving regulated parties less notice of prohibited or required conduct.”); Stephenson & Pogoriler, supra note 4, at 1461.
215 Paralyzed Veterans of Am. V. D.C. Arena L.P., 117 F.3d 579, 584 (D.C. Cir. 1997).
217 Id. at 256.
218 Id. at 257.
219 The cases were purposively chosen because they clearly identified both the relevant statute and the relevant regulation. I also sought a balance between cases that found parroting and cases that did not.
regulation for each measure by each group of cases. The positive numbers in this row for the non-parroting cases show that, in fact, the difference between the statute and the regulation in terms of the measures of opacity was positive, which means the opacity of the regulation went down on average. In contrast, the negative numbers in this row for the parroting cases show that, with the exception of the Flesch-Kincaid Grade Level scores, the regulation actually scored higher for opacity on average than the associated statute. Even on the Flesch-Kincaid Grade Level Score, the average difference for the parroting cases, while still positive, is lower than the associated increase in specificity for the non-parroting cases. It bears mentioning, as well, that the average differences on all four measures exceed the standard deviation for several cases.

### Table A.1: Assessing the Validity of the Measures Using Anti-Parroting Cases

<table>
<thead>
<tr>
<th></th>
<th>Word Count</th>
<th>F-K GL</th>
<th>Cog Complex</th>
<th>Obfuscation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Non-Parroting</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Haas</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Statute</td>
<td>0.243</td>
<td>3.879</td>
<td>-0.909</td>
<td>0.857</td>
</tr>
<tr>
<td>525 F.3d 1168</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regulation</td>
<td>0.201</td>
<td>0.979</td>
<td>1.643</td>
<td>0.005</td>
</tr>
<tr>
<td>( \Delta )</td>
<td>0.042</td>
<td>2.900</td>
<td>-2.552</td>
<td>0.852</td>
</tr>
<tr>
<td><strong>Villareal</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Statute</td>
<td>0.266</td>
<td>-1.693</td>
<td>-1.506</td>
<td>0.009</td>
</tr>
<tr>
<td>806 F.3d 1288</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regulation</td>
<td>0.207</td>
<td>1.464</td>
<td>-1.380</td>
<td>-0.362</td>
</tr>
<tr>
<td>( \Delta )</td>
<td>0.059</td>
<td>-3.157</td>
<td>-0.126</td>
<td>0.371</td>
</tr>
<tr>
<td><strong>Sierra Club</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Statute</td>
<td>0.240</td>
<td>0.042</td>
<td>-0.019</td>
<td>1.509</td>
</tr>
<tr>
<td>436 F.3d 1269</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regulation</td>
<td>0.266</td>
<td>-1.017</td>
<td>-0.089</td>
<td>-1.439</td>
</tr>
<tr>
<td>( \Delta )</td>
<td>-0.025</td>
<td>1.059</td>
<td>0.070</td>
<td>2.949</td>
</tr>
<tr>
<td><strong>Parroting</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Hagans</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Statute</td>
<td>0.201</td>
<td>-3.561</td>
<td>-1.383</td>
<td>1.613</td>
</tr>
<tr>
<td>694 F.3d 287</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regulation</td>
<td>0.203</td>
<td>-0.064</td>
<td>-0.524</td>
<td>-0.076</td>
</tr>
<tr>
<td>( \Delta )</td>
<td>-0.002</td>
<td>-3.497</td>
<td>-0.859</td>
<td>1.688</td>
</tr>
<tr>
<td><strong>N. Cal. River Watch</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Statute</td>
<td>0.216</td>
<td>-1.317</td>
<td>0.564</td>
<td>0.284</td>
</tr>
<tr>
<td>633 F.3d 766</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regulation</td>
<td>0.346</td>
<td>3.785</td>
<td>3.669</td>
<td>-1.409</td>
</tr>
<tr>
<td>( \Delta )</td>
<td>-0.130</td>
<td>-5.102</td>
<td>-3.105</td>
<td>1.693</td>
</tr>
<tr>
<td><strong>Gonzales</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Statute</td>
<td>0.161</td>
<td>-2.847</td>
<td>-0.589</td>
<td>-1.053</td>
</tr>
<tr>
<td>546 U.S. 243</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regulation</td>
<td>0.186</td>
<td>0.349</td>
<td>0.522</td>
<td>0.062</td>
</tr>
<tr>
<td>( \Delta )</td>
<td>-0.025</td>
<td>-3.196</td>
<td>-1.111</td>
<td>-1.115</td>
</tr>
<tr>
<td><strong>Average ( \Delta )</strong></td>
<td>0.025</td>
<td>1.391*</td>
<td>0.267</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Notes: Scores for F-K GL, Cog Complex, and Obfuscation are standardized z-scores. As in the body of the paper, higher scores for the statute and regulation cells indicate greater opacity; statistics in cells labeled with a “\( \Delta \)” always subtract regulation scores from statute scores, so positive values indicate that the regulation had lower opacity than the statute, while negative values indicate the opposite. A principal component analysis showed each of the four measures loading on one component with an Eigenvalue of 2.22. This one component accounted for 55.4 percent of the variation in the data. Values of “Average \( \Delta \)” that exceed the standard deviation for that variable for all cases are marked with an *.
To be sure, nothing definitive can be gleaned from a non-randomly selected sample of six cases, but the relatively consistent results in Table A.1 do provide some reassurance that the measures chosen for the study have some relationship to the underlying concept of interest.