

Copyright 2021 by Mitchell Chervu Johnston

Printed in U.S.A.
Vol. 116, No. 2

STEPIFICATION

Mitchell Chervu Johnston

ABSTRACT—Multistep tests pervade the law to the point that they appear to be a fundamental feature of legal reasoning. Famous doctrines such as *Chevron* or qualified immunity take this form, as do more obscure doctrinal formulas. But surprisingly, these doctrinal formulations *as a class* are relatively new. The reality is that the intellectual moment that gave rise to *Chevron* was one in which multiple older doctrines that relied on multifactor balancing were replaced by new tests formulated as multistep inquiries in which each step was a discrete inquiry.

This Article provides the first historical and normative account of this phenomenon—which I refer to as “stepification.” It charts both the rise of the new multistep tests as well as the intellectual climate that gave birth to these formulations, offering a theory of why courts chose to reorganize the law in this way at the time they did. Additionally, it argues that there are transsubstantive normative advantages and disadvantages to this mode of organizing doctrine, and it offers an accounting of the implications of historical stepification. In doing so, this Article aims to shed light on a historical phenomenon and on trends in modern legal disputes (such as recent cases over partisan gerrymandering and the future of *Auer*) that illustrate the work that stepification continues to do within our legal culture.

AUTHOR—Yale Law School, J.D. 2020. The views expressed in this Article are entirely my own and do not necessarily reflect the views of any other individual or organization. Primary thanks are owed to William Eskridge for his encouragement to write the paper that grew into this Article and to continue improving it as well as for his helpful comments on drafts. Thanks also to the participants in the Theories of Statutory Interpretation Seminar and participants in the Foundation of Legal Scholarship Seminar for valuable feedback. Thank you to Amy Chua, Michael Coenen, Jamie Durling, Kristen Houston, Peter Kallis, Dan Listwa, Michael Masciandaro, Sam Moyn, Nicholas Parrillo, Robert Post, David Pozen, Richard Re, Jed Rubinfeld, and Ravi Shah for helpful conversations, comments, and suggestions. Special thanks to Conor Dube, Abigail Fisch, Al Klevorick, Wajdi Mallat, and Simone Seiver for detailed comments and suggestions on drafts. Finally, thank you to the editors of the *Northwestern University Law Review* for their thoughtful suggestions and excellent editorial assistance.

NORTHWESTERN UNIVERSITY LAW REVIEW

INTRODUCTION 384

I. THE SPREAD OF STEPIFICATION 391

 A. *Defining Stepification* 391

 B. *A Threshold Objection: Just Words?* 394

II. STEPIFICATION OVER TIME 396

 A. *The Evolution of Stepification* 398

 B. *Stepification as an Intellectual Trend* 404

III. EXAMPLES OF STEPIFICATION (AND ITS DRAWBACKS) 416

 A. Chevron 416

 B. Lemon 420

 C. *Constitutional Criminal Procedure* 422

 D. *Partisan Gerrymandering* 425

IV. THE UNEASY CASE FOR (AND AGAINST) STEPIFICATION 428

 A. *The Case for Steps* 428

 B. *The Trouble with Stepification* 433

V. THE LIFE AND DEATH OF STEPIFICATIONS 447

 A. *The Life Cycle of Step Tests* 447

 B. *After Steps?* 453

CONCLUSION 454

APPENDIX: SAMPLE OF MULTISTEP TESTS 455

INTRODUCTION

Two Terms ago, in *Kisor v. Wilkie*, the Supreme Court decided whether to jettison the *Auer* deference doctrine.¹ *Auer* had answered the question of how to address agencies’ interpretations of their own regulations: first, a reviewing court was to ask whether the regulation is ambiguous, and second, if there is ambiguity, the court was to ask whether the interpretation is consistent with the regulation.² One might think the *Auer* test has a problem: it allows the agency to issue an ambiguous regulation and then to later clarify it via a less rigorous process.³ The petitioners in *Kisor* therefore asked the

¹ 139 S. Ct. 2400, 2408 (2019) (rejecting a challenge to the Court’s precedents, which held that courts should defer to an agency’s reasonable reading of its own genuinely ambiguous regulation).

² See *Auer v. Robbins*, 519 U.S. 452, 461 (1997). *Auer* does not precisely use this language, but this is essentially its test. *Auer* requires that the interpretation not be “plainly erroneous or inconsistent with the regulation.” *Id.* (quoting *Robertson v. Methow Valley Citizen’s Council*, 490 U.S. 332, 359 (1989)). This articulation in essence requires there to be room for interpretation—the regulation is ambiguous—and that the agency’s interpretation not be erroneous or inconsistent—that it is reasonable.

³ See *Decker v. Nw. Env’t Def. Ctr.*, 568 U.S. 597, 620 (2013) (Scalia, J., concurring in part and dissenting in part) (“But when an agency interprets its *own* rules—that is something else. Then the power

Court to replace *Auer* with the more amorphous *Skidmore* deference, which weighs the views of the agency based on a variety of factors.⁴ In short, they wanted to replace the two-step test, which they argued placed a “heavy weight on the scale[]” in favor of the agency, with a more flexible test.⁵ The Court, however, did no such thing. Instead, it added more steps; the new *Auer* test arguably has five steps.⁶ While the decision likely limits the reach of *Auer*, it is a triumph for the use of multistep tests to shape the law.

Multistep tests like *Auer* pervade the law. They structure decision procedures in administrative law,⁷ federal jurisdiction,⁸ the Establishment Clause,⁹ World Trade Organization disputes,¹⁰ and beyond.¹¹ Tests of this

to prescribe is augmented by the power to interpret; and the incentive is to speak vaguely and broadly, so as to retain a ‘flexibility’ that will enable ‘clarification’ with retroactive effect.”). For a recent review of the literature critiquing *Auer*, see generally Christopher J. Walker, *Attacking Auer and Chevron Deference: A Literature Review*, 16 GEO. J.L. & PUB. POL’Y 103, 105–10 (2018).

⁴ See Brief for Petitioner at 43, *Kisor*, 139 S. Ct. 2400 (No. 18-15); *Skidmore v. Swift & Co.*, 323 U.S. 134, 140 (1944) (“The weight of such a judgment in a particular case will depend upon the thoroughness evident in its consideration, the validity of its reasoning, its consistency with earlier and later pronouncements, and all those factors which give it power to persuade, if lacking power to control.”); see also *infra* Section III.A (discussing *Skidmore* in contrast with *Chevron* deference).

⁵ Brief for Petitioner, *supra* note 4, at 43.

⁶ See *Kisor*, 139 S. Ct. at 2415–17 (stating that *Auer* deference should only be afforded to interpretations of regulations when: (1) the regulation is genuinely ambiguous, (2) the interpretation is reasonable, (3) the interpretation is “one actually made by the agency,” (4) the interpretation actually implicates the agency’s “substantive expertise,” and (5) the “agency’s reading of [the] rule . . . reflect[s] ‘fair and considered judgment’” (quoting *Christopher v. SmithKline Beecham Corp.*, 567 U.S. 142, 155 (2012))); see also Christopher J. Walker, *What Kisor Means for the Future of Auer Deference: The New Five-Step Kisor Deference Doctrine*, YALE J. ON REGUL.: NOTICE & COMMENT (June 26, 2019), <https://yalejreg.com/nc/what-kisor-means-for-the-future-of-auer-deference-the-new-five-step-kisor-deference-doctrine> [<https://perma.cc/4EPX-7RU9>] (summarizing the *Kisor* decision). The Chief Justice’s concurrence reiterated this seemingly multistep test for applying *Auer*, stating that for deference to apply “[t]he underlying regulation must be genuinely ambiguous; the agency’s interpretation must be reasonable and must reflect its authoritative, expertise-based, and fair and considered judgment; and the agency must take account of reliance interests and avoid unfair surprise.” *Kisor*, 139 S. Ct. at 2424 (Roberts, C.J., concurring in part).

⁷ *Chevron U.S.A. Inc. v. Nat. Res. Def. Council, Inc.*, 467 U.S. 837, 842–43 (1984) (articulating a two-step test for deference to executive interpretation of the law).

⁸ *Gulf Offshore Co. v. Mobil Oil Corp.*, 453 U.S. 473, 477–78 (1981) (articulating a three-step test for rebutting the presumption of concurrent state jurisdiction); see also *Tafflin v. Levitt*, 493 U.S. 455, 459–60 (1990) (affirming the three-step test for rebutting the presumption of concurrent state jurisdiction in *Gulf Offshore*).

⁹ *Lemon v. Kurtzman*, 403 U.S. 602, 612–13 (1971) (articulating a three-prong test for determining violations of the Establishment Clause).

¹⁰ See Nita Ghei, *Evaluating the WTO’s Two Step Test for Environmental Measures Under Article XX*, 18 COLO. J. INT’L ENV’T L. & POL’Y 117, 119 (2007).

¹¹ See, e.g., *RJR Nabisco, Inc. v. European Cmty.*, 136 S. Ct. 2090, 2101 (2016) (articulating a two-step test for extraterritoriality); *Mayo Collaborative Servs. v. Prometheus Lab’ys, Inc.*, 566 U.S. 66, 76–80 (2012) (articulating a two-step test for determining whether an invention is sufficiently transformative

sort are so common that they can feel like a feature of law itself rather than a particular choice that has been made by judges to shape jurisprudence. Despite their apparent timelessness, however, the evolution of multistep tests and, in particular, the *rhetoric* of these tests is a relatively new and understudied phenomenon in American legal culture.¹²

Further, while the subject of multistep *procedural* processes has been studied by Professor Louis Kaplow,¹³ there has been little attention to the evolution of these sequential procedures within *doctrine* itself.¹⁴ While the logic of multistep doctrinal tests may resemble the best rationale for multistage adjudication—in the sense that both proceduralize decisions further—I believe the normative implications are importantly different.¹⁵

Many areas of law have undergone the *process* of what I call “stepification,” in which balancing tests are converted into multistep inquiries. In multistep inquiries, the legal decision-maker must answer a series of yes–no questions in which a “wrong” answer at any step terminates the inquiry. This is in contrast to tests that require balancing a set of factors or applying a standard—as opposed to a rule—to the conduct in question.¹⁶

so as to be a patentable application of natural laws or abstract ideas); *Batson v. Kentucky*, 476 U.S. 79, 80 (1986) (articulating a three-step inquiry for racial bias in peremptory strikes). These tests are not only created at the Supreme Court level. *See, e.g.*, *Disability Rights Mont. v. Batista*, 930 F.3d 1090, 1097 (9th Cir. 2019) (applying a two-prong inquiry to assess whether prison policies violate the Eighth Amendment). They are also sometimes recommended as a solution to issues that have at times vexed courts. *See, e.g.*, Brett M. Kavanaugh, *Fixing Statutory Interpretation*, 129 HARV. L. REV. 2118, 2144 (2016) (reviewing ROBERT A. KATZMANN, *JUDGING STATUTES* (2014)) (recommending a two-step procedure for statutory interpretation); Daniel B. Listwa, *Cooperative Covenants: Good Faith for the Alternative Entity*, 24 STAN. J.L. BUS. & FIN. 137, 159–170 (2019) (proposing a three-step framework for contract interpretation); Janice M. Mueller, *Determining the Scope of Copyright Protection for Computer/User Interfaces*, 9 COMPUTER/L.J. 37, 47 (1989) (proposing “a two-step test for infringement analysis in a computer/user interface copyright infringement case”); James Durling, Comment, *The Intercircuit Exclusionary Rule*, 128 YALE L.J. 231, 248 (2018) (recommending a three-step framework for addressing issues of intercircuit exclusion).

¹² *See infra* Section II.A (charting the rise of “stepified” language over time).

¹³ *See* Louis Kaplow, *Multistage Adjudication*, 126 HARV. L. REV. 1179, 1186–87 (2013) (studying the nature of procedural processes that test claims at multiple points—for example, with motions to dismiss and summary judgment—and providing a framework to evaluate such procedures).

¹⁴ In particular, Professor Kaplow’s work looks at procedures that test claims at multiple points in time with different substantive standards. My work here looks at how judges formulate sequential tests to evaluate one issue at one time.

¹⁵ In an important sense, both might flow from a common appreciation of a certain proceduralized aesthetic. Both seek to slice and dice some legal space in a sequential way via regulated decision points. In this sense, a connection might be drawn to Professor Pierre Schlag’s discussion of the “grid aesthetic” in American law. Pierre Schlag, *The Aesthetics of American Law*, 115 HARV. L. REV. 1047, 1051 (2002). Stepification, in my telling, is related to the same impulse that animates the desire for grids. *Id.* (“The grid aesthetic is the aesthetic of bright-line rules, absolutist approaches, and categorical definitions.”).

¹⁶ *See* Isaac Ehrlich & Richard A. Posner, *An Economic Analysis of Legal Rulemaking*, 3 J. LEGAL STUD. 257, 258–59 (1974) (explaining the distinction between rules and standards and how the distinction is somewhat artificial).

For example, the famous *Chevron* two-step test asks courts first to consider whether the statutory term is ambiguous.¹⁷ An answer of no results in a loss for the government. But, if the answer to the first question is yes, the court subsequently asks whether the government interpretation is reasonable.¹⁸ If the answer is no, the government loses at this second step. If the answer is yes, the government passes the test. Just like Professor Kaplow's structured procedures, each step is dispositive, but there is no final balancing inquiry.¹⁹ Thus, step tests take the form of a sequence of questions that are theoretically independent of one another. The answer to one question does not affect the analysis of the others, though it may obviate the need to answer further questions.

Whether a test has undergone stepification is a distinct question from whether the test takes the form of a rule or a standard.²⁰ As illustrated by *Chevron*, a two-step test can be two standards linked together. Neither of the *Chevron* steps is "rule-like" in practice. Therefore, to the extent that the step test converts the standards in question into a doctrinal approach that appears more rule-like, this effect is independent of whether the individual steps are actually rule-like themselves. But critically, for this reason, I do not want to make the claim that the rules versus standards debate is irrelevant to my inquiry.²¹ To the extent stepification makes doctrine more rule-like, the general arguments in favor of rules over standards will be closely related to the merits or drawbacks of the shift to multistep tests.

Furthermore, in a recent pair of insightful articles, Professor Kaplow has endeavored to study and critique what he terms "structured decision procedures," in which the court engages in a multistep inquiry rather than a

¹⁷ *Chevron U.S.A. Inc. v. Nat. Res. Def. Council, Inc.*, 467 U.S. 837, 842–43 (1984).

¹⁸ *Id.*

¹⁹ See Kaplow, *supra* note 13, at 1293.

²⁰ Briefly, standards are typically defined as open-ended directives that allow for later characterization of a variety of circumstances (for example, "drivers should drive at a reasonable speed given the circumstances"), whereas rules restrict the decision-maker significantly, perhaps to a single empirical fact (for example, "drivers should not exceed fifty-five miles per hour"). See Ehrlich & Posner, *supra* note 16, at 257. As Ehrlich and Posner acknowledge, this is not a sharp distinction but a matter of degree. *Id.* at 258.

²¹ See, e.g., PIERRE SCHLAG & AMY J. GRIFFIN, HOW TO DO THINGS WITH LEGAL DOCTRINE 100–18 (2020) (providing a critical overview of the distinction and an analysis of its recurrence within legal theory); Ehrlich & Posner, *supra* note 16, at 258, 261–280 (defining rules and standards and providing an economic analysis of the optimal choice for policymakers); Louis Kaplow, *Rules Versus Standards: An Economic Analysis*, 42 DUKE L.J. 557, 561–65 (1992) (providing an alternative overview of this debate); Duncan Kennedy, *Form and Substance in Private Law Adjudication*, 89 HARV. L. REV. 1685, 1687–1701 (1976) (discussing the jurisprudence of rules, "premised on the notion that the choice between standards and rules of different degrees of generality is significant"); Kathleen M. Sullivan, *The Supreme Court 1991 Term—Foreword: The Justices of Rules and Standards*, 106 HARV. L. REV. 22, 62–69 (1992) (discussing arguments for choosing a rule over a standard and vice versa).

balancing procedure.²² In particular, Professor Kaplow analyzes decision procedures that can be stylized in the following way: first, does the harm to the plaintiff exceed some threshold; second, if so, does the benefit of the action to the defendant equal some threshold; and finally, if steps one and two are satisfied, is the harm greater than the benefit.²³ He further notes that many tests take this exact form or a related form. My goal here is to expand on Professor Kaplow's observations in two ways. First, I aim to study a broader class of multistep tests, not all of which take the form that Professor Kaplow studies, such as *Chevron*. Second, when the focus is expanded, I find that a historical trend towards multistep tests can be observed, one perhaps driven by concerns with balancing summarized by Professor Kaplow.²⁴

This Article aims to provide the first comprehensive study of the process of stepification as well as its normative implications. In doing so, it makes two claims. The first is a descriptive and historical one, and the second is a normative commentary. The process of stepification, and the degree to which it has permeated U.S. law, is relatively new. In particular, the tendency of courts to speak of multistep tests seems to have begun in the 1970s and '80s, and the volume of cases using this language has grown steadily over time.²⁵ Although it is impossible to draw precise conclusions about the intentions of the disparate group of judges and lawmakers who built the stepified regime, evidence suggests that stepification should be understood as a response to discomfort with judicial discretion—a discomfort that resulted in a move to proceduralize the law. As Chief Justice John Roberts argued in *June Medical Services*, balancing can make equal treatment impossible because it asks judges to act as legislators.²⁶

Moreover, stepification emerged at the same time that several other formalistic trends among American elites took hold, offering clues about why this particular legal form emerged.²⁷ Stepification was a response to a desire among judges to portray the law as more formalistic and rule-bound

²² See Louis Kaplow, *Balancing Versus Structured Decision Procedures: Antitrust, Title VII Disparate Impact, and Constitutional Law Strict Scrutiny*, 167 U. PA. L. REV. 1375, 1376 (2019) [hereinafter Kaplow, *Balancing Versus Structured Decision Procedures*]; Louis Kaplow, *On the Design of Legal Rules: Balancing Versus Structured Decision Procedures*, 132 HARV. L. REV. 992, 993 (2019) [hereinafter Kaplow, *On the Design of Legal Rules*].

²³ Kaplow, *On the Design of Legal Rules*, *supra* note 22, at 993–94.

²⁴ *Id.* at 1047–55.

²⁵ See *infra* Section II.B.

²⁶ *June Med. Servs., L.L.C. v. Russo*, 140 S. Ct. 2103, 2136 (2020) (Roberts, C.J., concurring in the judgment).

²⁷ See *infra* Section II.B.2; cf. David E. Pozen & Adam M. Samaha, *Anti-Modalities*, 119 MICH. L. REV. 729, 781 (2021) (arguing that judges and lawyers exclude certain forms of argument as a way to maintain the legitimacy of the law as a discipline separate from politics).

than it may have previously seemed under regimes that favored multifactor balancing tests.

Thus, although much has been written about whether individual stepwise tests had an impact on the law and whether they remain relevant, these tests are best understood as a unified whole, part of a single historical trend rather than isolated changes in the law.²⁸ Regardless of the merits of *Chevron*, it is part of a broader trend in American jurisprudence.²⁹ This story has not yet been told in the legal literature. This Article aims to fill that gap. As a final observation, while this Article makes the claim that stepification as a historical phenomenon is an important and novel trend in several ways, my claim is not that the notion of steps in the law is itself new. Such a claim would almost certainly be false. For example, a negligence claim requires (1) a duty to the plaintiff, (2) a breach of that duty, (3) causation, and (4) a harm.³⁰

Parts I, II, and III chart the history of multistep tests. Part I offers a definition of stepification and distinguishes it from older legal forms that resemble multistep reasoning. Part II then examines trends in the language of stepification over time. It shows that courts have dramatically increased the degree to which they frame inquiries as stepified tests, which suggests a proliferation of these tests in important areas of law. Part II argues that this trend was driven both by desires to constrain judicial decision-making as well as by a broader movement among American elites towards game-theoretic and decision-theoretic analyses.³¹ Part III then discusses several case studies of the stepification process to show the universal qualities of these tests and their relevance to important areas of the law.

The second claim of the Article comes as a normative commentary on the character of stepified tests themselves. In Part IV, I argue that, seen as a group, these tests have both advantages and disadvantages as legal instruments. But while they may have their own individual dynamics and may be better suited to one doctrinal area or another, stepified tests have transsubstantive normative implications that apply broadly across their various manifestations.

²⁸ See, e.g., sources cited *infra* note 169 (differing over whether *Chevron* deference has actually altered the outcome of cases).

²⁹ For a discussion of the specific case of *Chevron*, see *infra* Section III.A.

³⁰ See, e.g., *Palsgraf v. Long Island R.R. Co.*, 162 N.E. 99, 99–100 (N.Y. 1928) (stating that a plaintiff must show that there was a duty to her as well a breach of that duty).

³¹ As I discuss in *infra* Section II.B, it is difficult to draw out the precise mix of these two factors, and therefore, without further empirical study, I am limited to providing suggestive evidence rather than concrete conclusions in this area.

I argue that stepification has important virtues. Stepified tests can increase the perception of the law as “lawlike” and hence can increase the perceived legitimacy of the judiciary. Similarly, each stepified test provides a structured framework for debate on a given topic, organizing legal argumentation as well as providing templates for decision-makers to explain their decisions. Unlike balancing tests, winners and losers in a stepified system can more easily understand the outcome even if they disagree with the conclusion. Finally, stepified tests can function as an effective means to organize and funnel lower court discretion and to make easy cases easy. They can provide clear instructions to lower courts on how to resolve cases and enhance predictability in certain circumstances.³²

That said, the implications of stepification are different from the merits and drawbacks of bright-line rules in some important ways. Multistep tests implicitly presume a particular normative structure that differs from that assumed by bright-line rules.³³ Further, multistep tests may conceal ambiguity that bright-line rules exist to strip out entirely. This might, I argue, make them inherently unstable in a particular way where that ambiguity requires ever more steps to be added to bring the doctrine into compliance with the appellate court’s view of the appropriate state of the law.³⁴ Again, I do not want to argue that a gulf necessarily exists—stepification may in fact help make a doctrine resemble a bright-line rule in some ways—only that there are important differences worthy of consideration.³⁵

Part IV also offers a commentary on the inherent instability of step tests and argues that because they flatten important normative issues, they must contort themselves over time to adjust to difficulties.³⁶ Although these issues can be relatively minor in some cases, in areas in which the normative stakes are higher, over-stepification can result in a breakdown of the legal regime in question and an end to the step test itself.³⁷ Part V then posits a tentative theory of the life and death of these tests over time. It argues that these tests’

³² See *infra* Section IV.A.

³³ See *infra* Section IV.B.1.

³⁴ See *infra* Part V.

³⁵ Cf. *infra* Section IV.A (arguing that multistep tests lead to a different pattern of precedent formation).

³⁶ In this sense, there may be a connection to Professor Frederick Schauer’s observation that hard rules “may have their edges rounded” over time as they are applied or interpreted via the creation of exceptions or selective overriding. See Frederick Schauer, *The Tyranny of Choice and the Rulification of Standards*, 14 J. CONTEMP. LEGAL ISSUES 803, 804–05 (2005). I am indebted to Michael Coenen for suggesting this article.

³⁷ Cf. Jeremy K. Kessler & David E. Pozen, *Working Themselves Impure: A Life Cycle Theory of Legal Theories*, 83 U. CHI. L. REV. 1819, 1821–22 (2016) (arguing that prescriptive legal theories in general often “work themselves impure” by becoming both more complicated and more compromised over time as a result of the normative challenges they encounter).

structure can be their undoing when subsequent cases put stress on the weaknesses their rigidity creates.

Ultimately, I do not declare definitively whether stepification is (and was) good or bad. Rather, the normative desirability of stepification is highly contextual. Just as there is no single answer to whether rules or standards are superior, or whether default rules should always be preferred to mandatory rules, there is no one answer to whether stepification is always positive or negative. Still, a common framework of *considerations* governs the evaluation of individual multistep tests; only once such jurisprudential moves are seen as a single trend rather than as individual doctrinal innovations do these dynamics come into view. The goal of this Article is to provide this framework to evaluate the merits of individual instances of stepification.

I. THE SPREAD OF STEPIFICATION

In some sense, the existence of multistep tests across the law should not be very surprising. Stepified tests are ways of codifying a decision-tree process. Therefore, it is not shocking that they exist in the law, which searches for ways to organize discretion and make decisions in a structured and consistent way.³⁸

But concluding that multistep tests have always been with us is a mistake. In particular, historical inquiry reveals that the language of stepification is a rather recent phenomenon, beginning in the 1970s and '80s. Moreover, stepification is a process that has occurred over time in various legal areas, converting previous modes of analysis into multistep tests. It can also include instances of courts adding additional steps to the inquiry in response to later imperfections in the test. This Part claims that when seen in proper context, stepification is a process that is worthy of study in its own right as a unified legal phenomenon, rather than as isolated incidents of doctrinal change.

This Part aims to make this claim in two Sections. The first Section works to define the notion of stepification and to clarify its relationship to the traditional debate over rules and standards. The second then addresses a threshold objection to my thesis—namely, that multistep tests have always been with us and so the change is illusory.

A. Defining Stepification

I use “stepification” to refer to a process by which the law is formalized into the language of multistep tests and has moved away from other forms of

³⁸ See *infra* Section IV.A.

inquiry. Thus, inquiries that have long been multistep—for example, asking whether the defendant committed the *actus reus* with the appropriate *mens rea*—do not fit within my thesis. In contrast, I focus on tests that have become stepified over time, such as the sliding-scale regime of *Skidmore* eventually becoming formalized into the *Chevron* two-step test, with further steps added over time.³⁹

For the purposes of this Article, I also mean to define a multistep test as one in which the decision-maker proceeds through a series of questions that are phrased as yes–no inquiries. Typically, at each stage, a particular answer can terminate the inquiry. For example, if the government fails at Step One of the *Chevron* analysis, then the inquiry is over.⁴⁰ Therefore, in a case involving a multistep test, the plaintiff or defendant can fail at any individual step of the inquiry.

That said, I do not want to entirely exclude from my focus what might be called “threshold” steps. While much of the argumentation here is focused on dispositive steps as part of answering a particular legal question, I also recognize that frequently, the language of steps is used to describe the question of whether or not to apply a given doctrine at all—for example, the so-called “Step Zero” of *Chevron*.⁴¹ In fact, implicit step tests are a constant in the law. For example, a court might ask first whether there is an actual contract at all before then asking whether the contract is unconscionable. Similarly, a court must first address whether something is speech before applying First Amendment protections.⁴² Because courts must always decide which regime applies before applying it—a step zero inquiry—there is always an implicit step test at work even if it is not styled as such. These steps, too, might have some of the merits and demerits discussed in Part IV insofar as they also represent on–off switches that presume a line can be drawn where one doctrine ends and the other begins.

Further, note that this definition does not require the court to explicitly use the language of “steps” for a test to take this form, though I use this

³⁹ See *infra* Section III.A.

⁴⁰ *Chevron U.S.A. Inc. v. Nat. Res. Def. Council, Inc.*, 467 U.S. 837, 842–43 (1984) (asking first whether the statute is ambiguous or silent on a given point).

⁴¹ Thomas W. Merrill & Kristen E. Hickman, *Chevron’s Domain*, 89 GEO. L.J. 833, 873 (2001) (coining the term “step zero” to describe “the inquiry that courts should undertake before moving on to step one of *Chevron*, or turning instead to *Skidmore* (or resolving the issue de novo)”).

⁴² See *Spence v. Washington*, 418 U.S. 405, 408–10 (1974) (engaging in the threshold inquiry of whether the individual’s communicative action fell “within the scope of the First and Fourteenth Amendments”). Similarly, a court must decide whether a “seizure” has occurred before analyzing whether it was “unreasonable” and thus violated the Fourth Amendment. See *Torres v. Madrid*, 141 S. Ct. 989, 1003 (2021) (holding that “the application of physical force to the body of a person with intent to restrain is a seizure even if the person does not submit and is not subdued” but noting that questions of reasonableness remain to be decided).

language as a proxy for the practice later and show that the use of such language has increased.⁴³ That said, how these tests are described is relevant. How the law is articulated is important even if the test is the same. Still, stepification can occur even if the word “step” is not initially used—though later courts may use this language. For example, the *Chevron* opinion itself does not use the word “step” or “prong” even as later commentators and judges have summarized it as creating a two-step test.⁴⁴

The form of the multistep test will sometimes vary depending on the work the steps do in each case. For example, if a step test requires that two conditions be met, then it may allow judges to dodge a question by finding that only one prong is relevant.⁴⁵ Alternatively, the inquiry can be more rigid, with one step necessarily following the other, as in *Chevron*, where it will frequently make little sense to resolve the second inquiry before the first.⁴⁶ This Article’s conclusions aim to discuss issues common to all multistep tests, but even within this family, there are important variations that will alter how my conclusions apply.

In sum, stepification is the conversion of either an open-ended standard or a multifactor balancing test into a multistep test, in which each step is potentially dispositive for the question. The individual steps need not have any strict sequencing, nor any relationship to previous steps. Furthermore, while the rhetoric of multistep tests matters, I do not require that the initial court that formulated the test refer to it explicitly in the language as steps, only that later interpreters do so consistently.

But the caveat that many legal inquiries seem to have a natural multistep structure raises a threshold objection to my thesis: Haven’t multistep tests always been with us? While the data presented later may demonstrate a rise in the language of multistep tests, this shift may only be rhetorical or, more directly, a distinction without a difference.⁴⁷ But I argue both that the multistep analyses found in private law have spread over time into public law and that the rhetorical shift from balancing to multistep tests is important, even if it ultimately crafts doctrine that reaches the same result. The next Section explicates these responses.

⁴³ See *infra* Section II.A.

⁴⁴ See 467 U.S. at 843–44.

⁴⁵ See, e.g., *Pearson v. Callahan*, 555 U.S. 223, 234 (2009) (holding that the mandatory two-step process announced in *Saucier v. Katz*, 533 U.S. 194 (2001), should be optional); see also *infra* Section II.B.

⁴⁶ Of course, an interpretation could be “reasonable” without a provision being ambiguous in theory. However, this circumstance is unlikely to generate litigation.

⁴⁷ See *infra* Section II.B (chronicling the rise in the number of cases using the language of multistep tests over time).

B. A Threshold Objection: Just Words?

Other legal inquiries present themselves as multistep inquiries as well. Again, tort analysis asks us to look for (1) a duty, (2) a breach of that duty, (3) causation, and (4) harm. Similarly, criminal law often asks us to look for (1) the *actus reus* and then (2) whether the act was taken with the requisite *mens rea*.⁴⁸ I do not deny (how could I?) that these are examples of multistep reasoning. Still, there are two reasons that this observation does not undermine my claim that stepification began in earnest in the 1970s and '80s and that its implications are normatively significant.

First, what sets stepification apart as a distinct, modern phenomenon is that it captures the migration of the traditional common law multistep inquiries into the realm of public law. As the data in the Appendix show, the growth in federal court discussion of multistep tests is not the result of an increase in tort analysis. Instead, legal issues such as personal jurisdiction and qualified immunity have acquired this structure. The notion of a multistep analysis has been moved over to these areas.

It would be one thing if multistep reasoning were somehow necessary to the analysis of these new areas, but it is not in two important ways. The first is that in many cases, the law *previously* lacked a multistep character.⁴⁹ Thus, in *Chevron*, the two-step analysis replaced an earlier balancing inquiry that looked to evaluate how convincing the agency's reasoning was.⁵⁰ Similarly, the defense of qualified immunity could have remained a defense of good faith. In *Pierson v. Ray*, the Supreme Court originally justified immunity based on the common law defense of "good faith and probable cause" in false arrest and imprisonment actions.⁵¹ It was only by departing from this formulation that the modern two-step test emerged.⁵²

Critically, the separation is *artificial* in an important sense, rather than a natural feature of the analytical problem. While the modern formulations of these doctrines can feel natural, it is by no means clear that they needed to be structured as multistep tests. In contrast, the criminal law's focus on both acts and mental states will all but require something that might resemble a two-step inquiry into each issue separately.⁵³ We know that the issue of

⁴⁸ See MODEL PENAL CODE §§ 2.01.1, 2.02.1 (AM. L. INST. 1962).

⁴⁹ See *infra* Part II (discussing trends in multistep tests over time); see also *infra* Part III (discussing specific examples).

⁵⁰ See *infra* Section III.A.

⁵¹ 386 U.S. 547, 557 (1967).

⁵² See *Harlow v. Fitzgerald*, 457 U.S. 800, 815–17 (1982).

⁵³ That said, while two steps are required for a finding of liability, degree does matter in criminal law. For example, which *mens rea* is shown will alter the nature of which homicide offense a defendant is guilty of. See MODEL PENAL CODE §§ 210.1–4.

deference to agency statutory interpretation does not require such a structure because it was not always so. This is not to say that the multistep version of these doctrines does not perform better or that their evolution in a stepwise direction was not in response to real deficiencies in the previous doctrinal regime.⁵⁴ But the existence of other nonstepified approaches demonstrates that the structure of the underlying issue does not directly require this approach. The move to make this separation is therefore normatively significant, even if other inquiries have this structure naturally.

A second point worth making here is that in many cases there is no need to even articulate the new doctrine in the language of steps. Professors Matthew Stephenson and Adrian Vermeule have, for example, convincingly argued that *Chevron* has only one step.⁵⁵ Similarly, qualified immunity could seemingly be framed as a one-step inquiry into whether an officer “violated clearly-established constitutional rights.”⁵⁶ While this is not true of all multistep tests, the ability to collapse some steps in this way suggests that in certain cases, the introduction of a multistep structure is a conscious *rhetorical* choice by courts rather than a change made out of necessity.

The first way to respond to the objection that stepification is nothing but a change in terminology is to observe that stepification captures a process by which courts adapt multistep thinking from other areas of the law and move it into new areas. The question then becomes why judges did so starting in the 1970s. As the data in Part II show, it is not as though courts immediately seized on multistep reasoning in the common law and used it to create multistep doctrines in other areas of the law. Before 1950, the language of multistep tests was essentially absent from case law.⁵⁷ Rather, the move to separate inquiries into multistep tests was responsive to a desire among lawyers to make the law more rule-like by adding analytical structure.⁵⁸ Stepification lent itself naturally to this process in cases where the law could not be neatly converted into bright-line rules. Even if stepification were equivalent to a one-step standard, it served an important rhetorical function.

This leads to the second response to the objection: the language of the law matters. That is, rhetorical shifts are important even if they have little

⁵⁴ For a discussion of the general benefits and costs of stepification, see *infra* Section IV.A.

⁵⁵ Matthew C. Stephenson & Adrian Vermeule, *Chevron Has Only One Step*, 95 VA. L. REV. 597, 598–600 (2009).

⁵⁶ This is contrary to the current formulation of qualified immunity that first asks whether a right has been violated, and then asks whether that right was clearly established. See *District of Columbia v. Wesby*, 138 S. Ct. 577, 589 (2018).

⁵⁷ See *infra* Section II.A.

⁵⁸ See *infra* Section II.B.

effect on whether Smith or Jones wins the case. The language judges use to justify their decisions affects the legitimacy of those decisions and how the audiences of opinions process the outcome. While a realist might point out that it likely makes little difference whether *Chevron* has one or two steps, that does not mean the form has no impact beyond the resolution of individual cases.⁵⁹

Additionally, the formalization of reasoning into something that has a name—a two-step test—may provide fertile ground for that form of labeling to grow. To the extent that a line can be drawn between different doctrines that unites them, that way of formulating the law may become more attractive to courts or lawyers looking to label their work.

To be sure, I believe that the use of the multistep form does have direct implications for how the judiciary operates.⁶⁰ In particular, I argue that the rigidity of multistep tests both makes certain decisions easier—an advantage—and creates problems elsewhere where the rigidity of the tests is overinclusive or underinclusive.

An observer should care about stepification because of its effect on the language of the law. In particular, the shift may suggest influences on the law and legal thinking from extralegal sources.⁶¹ The language of multistep tests may further influence how observers perceive the work of judges. Both are reasons to care about the trend, even if its impact on case-by-case adjudication is negligible. The next Part turns to providing a historical perspective on this trend.

II. STEPIFICATION OVER TIME

Stepification would be interesting as a matter of normative inquiry even if it had always been a feature of American law. Once stepification is seen in a broader historical context, however, it becomes more interesting as a jurisprudential matter. Indeed, looking at the data available, stepification, or at least the language of stepification, is a recent phenomenon in American law, not one with an extensive historical pedigree.

What is telling is that the number of cases using the specific language of steps has exploded over time. For example, between 1960 and 1970,⁶² thirty-six federal cases use language indicating the use of a two- or three-

⁵⁹ For an examination of *Chevron* in particular, see Richard M. Re, *Should Chevron Have Two Steps?*, 89 IND. L.J. 605, 608 (2014).

⁶⁰ See *infra* Part IV.

⁶¹ See *infra* Section II.B (discussing how stepification may be the result of external shifts in American culture).

⁶² The dates were filtered between January 1, 1960 and December 31, 1969.

step test.⁶³ In contrast, in the same period between 1990 and 2000 there were 5,380 cases using the same language. By the decade between 2000 and 2010, the number was 14,538. In total, a search of the Lexis database for terms associated with stepification reveals 61,247 federal cases that are framed as either two- or three-step tests.

While some of this growth can be attributed to famous two- and three-step tests that were articulated during this period, such as *Chevron*⁶⁴ and *Saucier*,⁶⁵ that does not explain the entire deviation.⁶⁶ The data in the Appendix support this conclusion, with only thirteen qualified immunity cases and three *Chevron* cases out of the random sample of 100 published cases.⁶⁷ Further, no alternative line of cases leaps out as being the lion's share of the published multistep cases. Instead, even after excluding multistep tests imposed by regulation, the random sample includes sixty-eight different doctrines taking a multistep form. It appears then that the process of stepification was the result of a general intellectual trend in the judiciary. Famous multistep tests were merely part of a broader shift.

Before moving on, I want to make a note on methodology. My goal for the various numbers in this Part (and in the Appendix) is to provide a suggestive quantitative picture of stepification. Given the large number of cases, it is impossible to verify that each case found via text searching actually applies a multistep analysis. In some cases, the judge may not say specific magic words, making the search underinclusive, and in others the court may apply a statutory test, making the search overinclusive.

After experimenting with search terms, I chose a term that seemed to capture a large share of tests without being overinclusive. Given that my goal is to demonstrate the rise of stepification, underinclusivity was preferable to a search that would pick up false positives. The data in the Appendix verify this understanding, with ninety-four of the 100 tests analyzed being examples of judge-made tests. Similarly, to avoid overinclusivity, I chose to exclude tests of four or more steps as they risked being either statutory or

⁶³ Searches were performed on Lexis due to its willingness to show the total search results even for large values. In particular, the search term used was "opinion((TWO or THREE) /1 (STEP or PRONG!) /1 (TEST or INQUIRY))." Results were then filtered by the date.

All searches, except those described in the Appendix, were performed on July 24, 2021. Because Lexis periodically adds to its database, readers attempting to replicate these searches may get slightly different numbers. Records of the searches are on file with the journal.

⁶⁴ See *infra* Section III.A.

⁶⁵ See *supra* note 45 and accompanying text. Note that while *Pearson* relaxed the mandatory order of the inquiry, it did not eliminate the two-prong inquiry.

⁶⁶ The difference is also surely due to the growth in the number of total cases. However, this cannot explain the deviation.

⁶⁷ Taking a 95% confidence interval for the total number of *Chevron* or qualified immunity cases in the total population gives a range of 7.9% to 22.4%.

regulatory creations or being balancing tests described in the language of multistep tests.⁶⁸ Finally, to verify that the data here are not an artifact of the growing number of cases, Figure 2 controls using a proxy for the number of cases in the database and shows that the trend is robust to restricting our attention to multistep tests as a fraction of cases.⁶⁹

The remainder of this Part aims to analyze this trend in two Sections. First, it looks at the history of the language of stepification over time, taking a high-level overview of the appearance of these terms in reported cases. It then attempts to provide an explanation for these trends by examining potential internal and external stories that tell why stepification arose at the time it did. Still, the goal of this Part is not to offer a definitive historical or empirical analysis of this trend. Rather, my claim is limited to the idea that, whatever implicit steps have always existed in the law, the rise of the language of stepification is a recent trend, not a constant of legal analysis. That is to say, while multistep tests may seem to be omnipresent, they are in fact a creation of the current legal moment.

A. The Evolution of Stepification

Upon looking at the data, what is striking is the shocking absence of multistep or multiprong language in the earlier parts of legal history. A search of the Lexis database for terms associated with stepification reveals 61,247 federal cases (14,568 published) that frame their inquiries as two- or three-step tests.⁷⁰ However, limiting the search to the period from 1875 to 1950 reveals only one case, state or federal, that uses this terminology.⁷¹ The remainder of this Section proceeds to evaluate this shift in the number of

⁶⁸ Moreover, adding a greater number of steps to the analysis did not change the result much. When the search string was changed to “opinion((TWO or THREE or FOUR) /1 (STEP or PRONG!) /1 (TEST or INQUIRY)),” the total number increased to only 64,022 cases in Lexis.

⁶⁹ Based on the helpful advice of the Yale Law School research librarians, I looked for cases with at least four uses of the word “the.” The goal was to exclude orders that might be coded as cases in the database (for example, grants of the writ of certiorari) while still capturing the majority of cases.

⁷⁰ Returning again to the search term of “opinion((TWO or THREE) /1 (STEP or PRONG!) /1 (TEST or INQUIRY))” for these searches.

⁷¹ *Universal Form Clamp Co. v. Taxis*, 267 F. 578, 580 (7th Cir. 1920) (“To compare appellant’s present with its former device to determine equivalency would be to substitute for the elements of the claim as written the mechanical equivalents of appellant’s first structure. It would be giving patentee a double or two-step test for determining equivalency.”). Specifically, the dates used were January 1, 1875 and December 31, 1949. The first date was chosen because this is the first case that is mentioned in the database. Relaxing the search term to “opinion((TWO or THREE) /2 (STEP or PRONG!) /2 (TEST or INQUIRY))” adds a second case. *Adam v. Conn. Med. Examining Bd.*, 16 Conn. Supp. 281, 283 (Super. Ct. 1949) (“Three steps of inquiry are indicated: (1) What were the charges? (2) What act does the statute provide for revocation? (3) On the findings made upon the evidence, are the conclusions justified in the light of the statute?”). Including the possibility of four-step tests using the search string “opinion((TWO or THREE or FOUR) /2 (STEP or PRONG!) /2 (TEST or INQUIRY))” does not further alter the results.

cases by examining three different time periods to track the evolution of this trend over time.

Further, one point of this Section is to observe that the language of these tests did not exist in the way it does now, despite the fact that similar tests may have existed in the past. Moreover, examination of individual cases reveals that the tests that did exist related more to instances of multiprong offenses rather than replacements for multifactor balancing tests. The trend observed using this method is suggestive of a true shift in the underlying legal culture rather than merely a shift in terminology.

1. 1875 to 1950

Just one federal or state case employs the language of stepification in this time period. But, while this case actually uses the terms in question, it does not use them to articulate a doctrinal test. In *Universal Form Clamp Co. v. Taxis*, the Seventh Circuit addressed a patent infringement claim.⁷² While the opinion uses the phrase “two-step test,” the reference is not to an existing doctrinal test but rather to a two-step test proposed by the appellees, which the court rejects.⁷³ A look at the history of this period more generally, however, reveals that multistep frameworks did exist, though they were often not couched in this language.

For example, running the original search but including summaries⁷⁴ highlights a second federal case during this time period. In *Aetna Life Insurance Co. v. France*, the Supreme Court considered a claim to recover on a life insurance policy.⁷⁵ The Court noted that the jury had to answer two questions to evaluate the case: “first, [w]as the representation made? second, [w]as it false?”⁷⁶ This latter case illustrates the point made above: the law has never been entirely without steps. But their prevalence has grown and the language used to describe these doctrines has shifted—perhaps as the formal notion of a multistep test took form.

My claim here is not that there were no multiprong tests during this time period—there were—but rather that courts were seemingly not creating new tests nor speaking in the language of multistep tests.⁷⁷ Moreover, the data in

⁷² 267 F. at 580.

⁷³ *Id.* (rejecting appellee’s argument that because appellant’s first device was admitted to be an infringement and because the “second structure is the mechanical equivalent of the first, infringement by appellant’s second structure is shown” as an impermissible “two-step test for determining equivalency”).

⁷⁴ This means that the term searched was: “(TWO or THREE) /1 (STEP or PRONG!) /1 (TEST or INQUIRY).” This allows the search to include results from the summary portion of the Lexis entry that do not appear in the opinion.

⁷⁵ 91 U.S. 510, 510 (1876).

⁷⁶ *Id.* at 516.

⁷⁷ See *supra* Section I.B.

the Appendix suggest that modern multistep tests are not simply restatements of older tests reexpressed in new language. Rather, *Chevron* potentially adopted the current two-step language because it was born in a fertile moment for that sort of rhetoric. Further, around this time, more and more areas that did not naturally break into several component parts were nevertheless converted into multistep tests.⁷⁸

2. 1960 to 1970

Moving forward to the 1960s reveals a small uptick in cases and a general trend towards modern stepification. While the number of reported cases remains small⁷⁹—thirty-three federal⁸⁰ and eighteen state⁸¹—many of these cases apply multistep tests to resolve more complicated issues than those presented in cases from earlier times. We see here an indication of the development of stepification, as well as evidence that its emergence as a widespread trend postdates this period.

The cases over this period evince the development of multistep tests to evaluate certain normative questions. For example, in *United States ex rel. Thurmond v. Mancusi*, the district court, without citation, engaged in a “two step inquiry” to determine whether a guilty plea was voluntary.⁸² Similarly, in *King v. Gardner*, the Fifth Circuit drew on Fourth Circuit law that had developed a two-step test to determine whether a disability existed.⁸³

During this time, district courts also described doctrine derived from other opinions in the language of steps. For example, in *Wirtz v. Monarch Patrol & Detective Service, Inc.*, the district court applied a two-step test it derived from Supreme Court precedent to determine whether a given activity or service was interstate commerce.⁸⁴ The court acknowledged that it was drawing its test from earlier cases but chose to summarize the doctrine in the

⁷⁸ See *infra* Part III (discussing examples).

⁷⁹ There are three unreported federal cases and zero unreported state cases.

⁸⁰ But note that six of these federal cases refer to multistep tests in the context of medical tests as opposed to doctrinal tests. See *Mark v. Celebrezze*, 348 F.2d 289, 293 (9th Cir. 1965); *McGiven v. United States*, 183 Ct. Cl. 920, 928 (1968); *Morgan v. Gardner*, 254 F. Supp. 977, 980 (N.D. Okla. 1966); *Riddle v. Celebrezze*, 235 F. Supp. 657, 660 (W.D.S.C. 1964); *McCord v. Celebrezze*, 221 F. Supp. 206, 209 (W.D.S.C. 1963); *Ferrell v. Gardner*, 260 F. Supp. 996, 1000 (S.D. W. Va. 1966).

⁸¹ The same search term was used on Lexis for this result. Specifically, the term searched was “opinion((TWO or THREE)/1 (STEP or PRONG!)/1 (TEST or INQUIRY)).” The dates in question were January 1, 1960 to December 31, 1969.

⁸² 275 F. Supp. 508, 515 (E.D.N.Y. 1967).

⁸³ 391 F.2d 401, 403 (5th Cir. 1967) (“It is clear from the statutory language that the determination of disability involves a two-step inquiry.” (citing *Cyrus v. Celebrezze*, 341 F.2d 192, 194 (4th Cir. 1965)); see also *Thomas v. Celebrezze*, 331 F.2d 541, 545 (4th Cir. 1964) (articulating first the test from *Cyrus* by noting that “[t]here really are two steps to a finding of disability”).

⁸⁴ 1967 U.S. Dist. LEXIS 7910, at *2–3 (E.D. Pa. Aug. 18, 1967) (citing *Durkin v. Joyce Agency, Inc.*, 110 F. Supp. 918 (N.D. Ill. 1953), *aff’d*, 348 U.S. 945 (1955)).

language of steps, arguing that these previous cases “essentially applied [a] two step test.”⁸⁵

It is difficult to draw firm conclusions from these scattered cases, but two general lessons stand out. First, there was an increased use of the language of multistep tests to describe what the courts in question were doing. The main judicial test being applied by the federal courts seems to be the *Aguilar* two-step test to evaluate the reliability of search warrants based on anonymous information.⁸⁶ Other tests, however, were also being used.⁸⁷ Second, the number of cases that used this language was still remarkably small. While the evidence indicates that multistep tests were in use at the time in multiple areas of the law,⁸⁸ the fact that there are relatively few examples of these cases suggests that the number of multistep tests being applied at the appellate level was still fairly small. Indeed, relaxing the search term to look for only examples of the use of phrases like “two step” and “three prong” produces only 699 results over the same time period.⁸⁹ This further evidences that judges at this time were thinking in the language of steps in general, though the terminology was still emerging during this time.

3. 2000 to 2010

Thirty years later, the trend has accelerated.⁹⁰ The original search produces 14,538 cases using the language of stepification, 4,093 of which are published. Moreover, these cases cover a broad range of doctrinal areas from the more familiar, such as qualified immunity,⁹¹ to the less well known, such as equal educational opportunity.⁹²

⁸⁵ *Id.*

⁸⁶ For examples of federal courts utilizing the *Aguilar* two-step test, see *Spinelli v. United States*, 393 U.S. 410, 413 (1969), and *McCreary v. Sigler*, 406 F.2d 1264, 1268 (8th Cir. 1969).

⁸⁷ See, e.g., *D/S A/S Flint v. Sabre Shipping Corp.*, 228 F. Supp. 384, 387 (E.D.N.Y. 1964) (applying a two-step test to resolve a service of process issue in admiralty cases).

⁸⁸ See, e.g., *Printing Plate Supply Co. v. Crescent Engraving Co.*, 246 F. Supp. 654, 660–61 (W.D. Mich. 1965) (applying a “three-step test of patentability”).

⁸⁹ Specifically, this result was obtained looking over the January 1, 1960 to December 31, 1969 period using the search term “opinion((TWO or THREE) /1 (STEP or PRONG!)).”

This modification was chosen to capture instances where the courts expressed a decision as a two-step process without categorizing it as a two-step doctrinal test. Additionally, a look at the search results indicates that the 699 number may be high, as many results refer to steps in reference to the facts of the case as opposed to the doctrine being applied.

⁹⁰ Of course, one could attempt to tell the story of every intervening decade as well. My point of skipping ahead is to just illustrate the current status quo as stepification has taken hold.

⁹¹ For an example of a lower court considering the qualified immunity defense during this time period, see *Chappell v. City of Cleveland*, 584 F. Supp. 2d 974, 988–99 (N.D. Ohio 2008).

⁹² See, e.g., *United States v. Texas*, 572 F. Supp. 2d 726, 732 (E.D. Tex. 2008) (citing *Castaneda v. Pickard*, 648 F.2d 989 (5th Cir. 1981)) (articulating a three-prong test for compliance with 20 U.S.C. § 1703(f)), *rev'd*, 601 F.3d 354 (5th Cir. 2010).

Indeed, a look at the six Supreme Court cases⁹³ that explicitly employ the language of step tests during this period illustrates these tests' broad application, spanning a variety of different doctrinal areas ranging from reservation of land⁹⁴ to obscenity.⁹⁵ Relaxing the search term to include any cases in which the Court mentions steps at all reveals additional cases in which the Court is working within a stepified framework or discussing an existing doctrinal test.⁹⁶ Though this sample undoubtedly fails to capture the full extent to which the Court is engaging in stepwise reasoning, it points to the expansion of this language and to the fact that the growth in multistep cases was not limited to a few famous examples.

This historical evolution has two general implications. First, there is a change in the usage of the language of steps. If there is one thing that this method can establish, it is that courts and judges have taken to using the language of steps more broadly. This trend has occurred not only in the famous areas discussed above, but also in a variety of other doctrinal areas.⁹⁷

Second, and more tentatively, the data suggest that there was an increase in the number of step tests being used doctrinally. While the information from earlier eras suggests that step tests have always been a legal reality, the dramatic increase in the language of multiprong tests suggests a broader trend. The evidence grows stronger still when coupled with the qualitative examples discussed earlier of courts and scholars converting areas of law into these new stepwise regimes over time.

* * *

As noted repeatedly throughout this Section, it is difficult to make a definitive claim about exactly what happened to cause a shift to occur suddenly in the 1970s and '80s. There were relatively few cases using the language of stepification before 1970. The number then continued to grow slowly during the '70s before a period of steady growth began in the '80s

⁹³ Relaxing the search term to "opinion((TWO or THREE) /1 (STEP or PRONG!))" alters this number to forty-six cases at the Supreme Court level.

⁹⁴ *Alaska v. United States*, 545 U.S. 75, 100 (2005); *Idaho v. United States*, 533 U.S. 262, 273–74 (2001).

⁹⁵ *Ashcroft v. ACLU*, 535 U.S. 564, 574–79 (2002) (applying the three-prong *Miller* test). The other cases also concern a variety of issues. See *Pearson v. Callahan*, 555 U.S. 223, 232 (2009) (considering qualified immunity defense); *Rice v. Collins*, 546 U.S. 333, 338 (2006) (applying *Batson* three-step framework).

⁹⁶ See, e.g., *Nat'l Cable & Telecomms. Ass'n v. Brand X Internet Servs.*, 545 U.S. 967, 986 (2005) (discussing the *Chevron* framework); *Barhart v. Thomas*, 540 U.S. 20, 24–25 (2003) (applying the Social Security Administration's five-step evaluation to determine disability).

⁹⁷ See *infra* Part III.

that has not yet abated. The same trend, albeit more drastic in terms of the absolute number of cases, can be seen in unpublished opinions. Both trends are illustrated in Figures 1A and 1B, respectively.

FIGURE 1A: REPORTED FEDERAL CASES USING THE LANGUAGE OF STEP TESTS⁹⁸

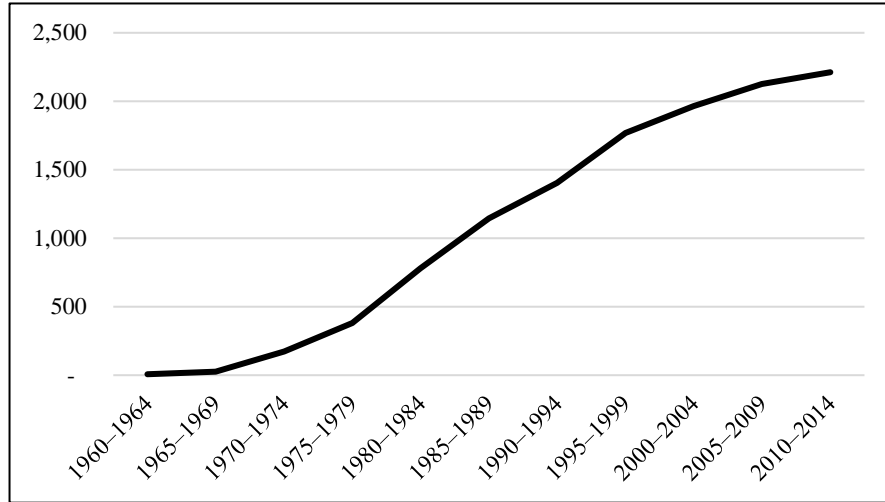
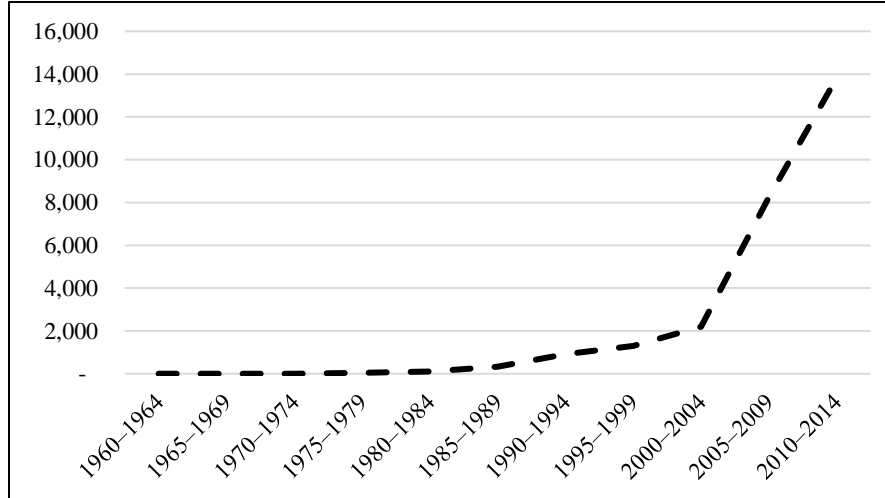


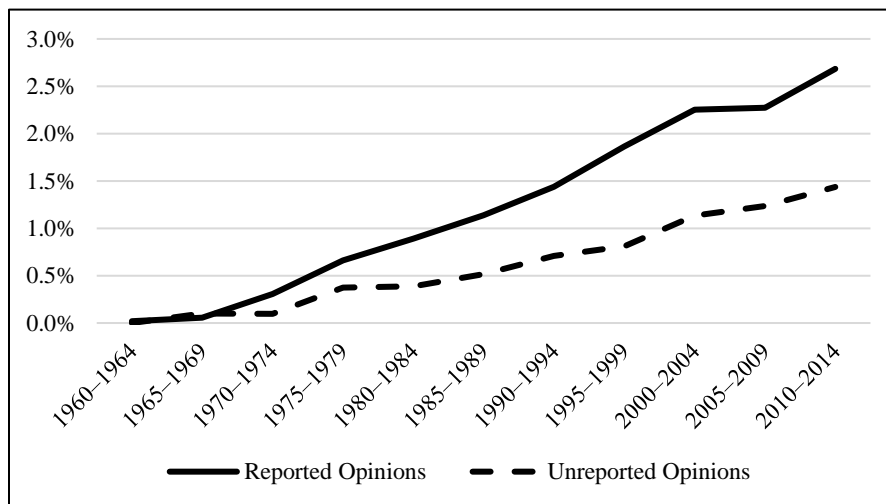
FIGURE 1B: UNREPORTED FEDERAL CASES USING THE LANGUAGE OF STEP TESTS



⁹⁸ Cases were looked at in five-year increments. The search term used was “opinion((TWO or THREE) /1 (STEP or PRONG!) /1 (TEST or INQUIRY)).” A similar trend can be observed in the relevant state cases. Initially, more reported cases used the language of stepification, but the number of unreported cases grew more quickly than reported cases. From 2005 to 2009 there were approximately 2,700 cases of each type, but from 2010 to 2014 there were over 4,000 unreported cases and fewer than 3,000 reported cases.

The trend is also robust to controlling for the total number of reported and unreported cases. The proportions of reported and unreported federal cases using the language of steps are depicted in Figure 2:

FIGURE 2: PERCENTAGE OF FEDERAL CASES USING THE LANGUAGE OF STEP TESTS⁹⁹



Analysis of the cases seems to preclude an argument that one or two doctrinal areas drove all of this growth. And indeed, the random sample included in the Appendix of 100 cases from between 2010 and 2014 reinforces this conclusion, with only thirteen qualified immunity cases and three *Chevron* cases. At some point, judges started to turn to the language of steps to express their ideas. In some cases, this language merely codified the previous status quo, but in others it transformed the doctrine into its modern form, replacing more amorphous processes with more regimented ones. Exploring the reasons for this shift is the project of the next Section.

B. Stepification as an Intellectual Trend

The evidence of stepification suggests a trend in American jurisprudence that has previously gone unnoticed. While from our modern vantage point, steps may seem a mundane feature of legal decision-making, the extent of their use is a novel phenomenon and the language of multistep tests is the result of an intellectual shift among judges.

The examination of the cases indicates, however, that no “smoking gun” explanation for the growth of stepification suggests itself. For example,

⁹⁹ Again, the total number of reported cases was determined by searching “opinion(atleast4(THE)).” See *supra* note 63 (describing the reasoning for this search).

rather than a single case driving the trend, the rise of steps occurs across different areas seemingly at once.

Despite this, we can speculate about the causes of this shift. The remainder of this Section proceeds by first offering an internal hypothesis for the trend, one driven by lawyers and judges. It then offers a second, more external perspective of stepification as part of a broader movement among American elites towards decision theory as an approach to managing complexity.

1. *Internal Trends*

One story that could be told is that the trend towards multistep tests is a part of the broader movement in U.S. law towards proceduralization and away from balancing. It was part of a move to focus on procedure and occurred contemporaneously with the rise of textualist and originalist methodologies that claimed, in part, to be justified by their rule-like nature as opposed to more subjective inquiries.¹⁰⁰ Professor Kaplow, for example, notes that there has long been a “queasiness” about balancing and a desire to disguise its operation using structure.¹⁰¹ And indeed, in his dissent in *June Medical Services L.L.C. v. Russo*, Justice Neil Gorsuch argued explicitly that interest-balancing was inappropriate for courts, whose work should be restricted to structured inquiries that avoid evaluation of “competing social interests.”¹⁰²

Justice Gorsuch’s position fits well within the post-New Deal history of the law. Professor Thomas Grey has written that after the New Deal there was a doctrinal turn towards procedure as opposed to substantive law.¹⁰³ While this is historically earlier than stepification, there is further evidence to suggest that the development of the legal process school reoriented the law towards procedure.¹⁰⁴ The map is not perfect, as stepification is not

¹⁰⁰ See, e.g., AMANDA HOLLIS-BRUSKY, *IDEAS WITH CONSEQUENCES: THE FEDERALIST SOCIETY AND THE CONSERVATIVE COUNTERREVOLUTION* 1–3 (2015) (dating the origins of the conservative legal movement to the start of the Reagan Administration in 1980); KEN I. KERSCH, *CONSERVATIVES AND THE CONSTITUTION*, at viii (2019) (arguing that the roots of the conservative legal movement grew out of Richard Nixon’s victory in 1968 and triumphed with the election of Ronald Reagan).

¹⁰¹ Kaplow, *On the Design of Legal Rules*, *supra* note 22, at 1047–48.

¹⁰² 140 S. Ct. 2103, 2179 (2020) (Gorsuch, J., dissenting).

¹⁰³ See Thomas C. Grey, *Modern American Legal Thought*, 106 *YALE L.J.* 493, 502 (1996) (reviewing NEIL DUXBURY, *PATTERNS OF AMERICAN JURISPRUDENCE* (1995)) (“Three elements defined the Process approach: focus on the rule of law as a value essential to the preservation of liberal democracy; support for the New Deal and the modern administrative and welfare state; and doctrinal emphasis jurisdiction and procedure as against substantive law.”).

¹⁰⁴ William N. Eskridge, Jr. & Gary Peller, *The New Public Law Movement: Moderation as a Postmodern Cultural Form*, 89 *MICH. L. REV.* 707, 722–23 (1991) (“The normativity of law rests in its process, and not in its substance. As we shall now see, however, legal process’ proceduralism was itself

necessarily proceduralization, but the analogy is suggestive. Stepification proceduralizes a particular doctrinal issue of the law.¹⁰⁵ In particular, the proceduralization of stepification adds rigidity to the process by which a decision is reached, allowing the result of the case to be articulated as a set of answers to disjoint questions.

To put the point more sharply, stepification can be seen as formalism without ideal forms. It works to formalize jurisprudence without the need for a Langdellian view that a number of underlying principles animate the law. Professor Langdell argued that lawyers could use scientific methods to derive correct legal judgments from a few basic principles and concepts.¹⁰⁶ This feature of stepification is demonstrated by the large number of areas of law to which it has been applied without any apparent thought given to the underlying topic.¹⁰⁷

It is then notable that stepification occurred after a period in which there was a move towards balancing and away from bright-line rules. As Professor John Langbein and his coauthors have highlighted, the original *Restatements* appeared between 1932 and 1944 and were immediately critiqued by the realists as insufficiently nuanced and overly simplistic.¹⁰⁸ This dynamic may have been reflective of the general trend of the codification movement: Professor Grant Gilmore makes what is perhaps a parallel critique of Professor Langdell's early casebooks by observing that Professor Langdell seemingly pruned out nonconforming cases in his mission to present the evolution of the law towards "the ideal of the one true rule of law."¹⁰⁹ In response to this challenge, Professor Langbein observes that there was a movement within the *Second Restatements* towards a balancing approach that would capture the nuance that the realists charged was lacking.¹¹⁰ The period before stepification was therefore marked by the rise of balancing to express the nuance the realists perceived was missing from the formalist picture of the law.

subject to different ideological interpretations, and almost immediately in history dissolved into competing normative visions.").

¹⁰⁵ See *infra* Part III.

¹⁰⁶ Thomas C. Grey, *Langdell's Orthodoxy*, 45 U. PITT. L. REV. 1, 5 (1983) ("The heart of the theory was the view that law is a science. Langdell believed that through scientific methods lawyers could derive correct legal judgments from a few fundamental principles and concepts, which it was the task of the scholar-scientist like himself to discover.").

¹⁰⁷ See *supra* notes 7–10 and accompanying text.

¹⁰⁸ See JOHN H. LANGBEIN, RENÉE LETTOW LERNER & BRUCE P. SMITH, *HISTORY OF THE COMMON LAW: THE DEVELOPMENT OF ANGLO-AMERICAN LEGAL INSTITUTIONS* 853–54 (2009). I am indebted to Michael Masciandro for directing me towards this point.

¹⁰⁹ GRANT GILMORE, *THE AGES OF AMERICAN LAW* 42–43 (2d ed. 2014).

¹¹⁰ LANGBEIN ET AL., *supra* note 108, at 853–54.

But the embrace of balancing as a way to capture and express complexity did not last. In particular, the other historical event lurking in the background is the broader legacy of the Warren Court. This is not to say that stepification is a response to a particular Warren Court decision or set of decisions. Rather, another suggestive connection is that the worry about judicial discretion was transformed into a move to proceduralize the law. Of course, as I argue below, this transformation was not an escape from substantive law. Rather, as Professor Richard Re points out, steps themselves have normative valence even if equivalent logical formulations with greater or fewer steps exist.¹¹¹ The rhetoric of stepification, however, is a refuge from certain accusations of discretion.¹¹² Stepification offers, at the very least, the appearance of order. This is not to say that it does not have benefits, but its power is tangled up in its ability to give the work of judges a certain appearance of neutrality. Moreover, it can provide appellate courts with the ability to control lower court discretion via the creation of more rule-like structures to guide decision-making.¹¹³

Stepification, then, arose at a time when at least some observers were concerned about judicial discretion. Indeed, several parallel shifts explicitly in response to the Warren Court similarly looked to limit the discretion of judges. These shifts focused the inquiry in cases on single questions rather than on more complicated multifactor approaches. For example, the textualist movement was driven, in part, by an effort to simplify the process of interpretation and to make it more predictable.¹¹⁴ While textualism has certainly been defended on other grounds,¹¹⁵ its appeal to a single question—what do the words mean—has given it a rhetorical advantage over

¹¹¹ See Re, *supra* note 59, at 608.

¹¹² See Pozen & Samaha, *supra* note 27, at 741–68 (arguing that constitutional lawyers must exclude certain modes of reasoning to stabilize the distinction between law and politics).

¹¹³ See *infra* Section IV.B.

¹¹⁴ See, e.g., Frank H. Easterbrook, *Statutes' Domain*, 50 U. CHI. L. REV. 533, 544–47 (1983) (defending a textualist approach in part based on the fact that it “gives the legislature a low-cost method to signal its favored judicial approach to public interest legislation” and “recognizes that courts cannot reconstruct an original meaning because there is none to find”); see also William N. Eskridge, Jr. & Philip P. Frickey, *Statutory Interpretation as Practical Reasoning*, 42 STAN. L. REV. 321, 340–41 (1990) (“By emphasizing the statutory words chosen by the legislature, rather than (what seem to be) more abstract and judicially malleable interpretive sources, textualism also appeals to the values of legislative supremacy and judicial restraint.”).

¹¹⁵ See, e.g., *K Mart Corp. v. Cartier, Inc.*, 486 U.S. 281, 323–25 (1988) (Scalia, J., concurring in part and dissenting in part) (arguing that democracy requires judges to rely on the text of the statute); John F. Manning, *What Divides Textualists from Purposivists*, 106 COLUM. L. REV. 70, 73 (2006) (“[T]he ‘new textualism’ challenged the prevailing judicial orthodoxy by arguing that the Constitution, properly understood, requires judges to treat the clear import of an enacted text as conclusive . . .”).

competitors that focus on more ambiguous inquiries.¹¹⁶ Professor Caleb Nelson has observed as much, noting that textualists have an affinity for rule-like methods of interpretation over standards.¹¹⁷ What the text means is not as sharp and crisp as the typical legal bright line (though it may be an improvement); the issue is whether something is rule-like in nature. Textualism, especially its strictest version, which rejects all other sources, offers just such a rule-like interpretive process.¹¹⁸ Textualism presents itself as a highly proceduralized project that aspires to neutrality and objectivity. Its tools, including a list of Latin canons, give the appearance of a largely mechanical process of interpretation, taking the judge out of the equation.

Similarly, originalism shares textualism's goal of making interpretation more predictable and less personal.¹¹⁹ Indeed, originalism itself is a sort of stepified constitutional inquiry, though it is often not expressed this way and various originalist theories might operate differently. It requires the interpreter to look first to the text of the Constitution and its original meaning and then, if the text is unclear, to structure and history for insight.¹²⁰ If these sources are ambiguous, then other sources may come into play.

Originalism can perhaps be seen as a stepification of Professor Phillip Bobbitt's famous six modalities of constitutional interpretation.¹²¹ First, at the "interpretation" stage, text, history, and structure can be used to discover the linguistic meaning of the provisions of the Constitution.¹²² Second, construction may be required to give legal effect to the linguistic meaning,

¹¹⁶ See Easterbrook, *supra* note 114, at 533 ("The construction of an ambiguous document is a work of judicial creation or re-creation. Using the available hints and tools—the words and structure of the statute, the subject matter and general policy of the enactment, the legislative history, the lobbying positions of interest groups, and the temper of the times—judges try to determine how the Congress that enacted the statute either actually resolved or would have resolved a particular issue if it had faced and settled it explicitly at the time.")

¹¹⁷ Caleb Nelson, *What Is Textualism?*, 91 VA. L. REV. 347, 350 (2005).

¹¹⁸ Eskridge & Frickey, *supra* note 114, at 340–41 (describing the two strains of textualism).

¹¹⁹ See, e.g., Mary Wood, *Scalia Defends Originalism as Best Methodology for Judging Law*, UNIV. OF VA. SCH. OF L. (Apr. 20, 2010), https://www.law.virginia.edu/news/2010_spr/scalia.htm [<https://perma.cc/34RL-JH3K>] (reporting on a speech by the late Justice in which he argued that originalism properly restricted the discretion of judges).

¹²⁰ John O. McGinnis & Michael B. Rappaport, *Symmetric Entrenchment: A Constitutional and Normative Theory*, 89 VA. L. REV. 385, 391 (2003) ("Under this approach, one first examines the meaning of the Constitution from an eighteenth-century perspective to determine whether it unambiguously resolves the question. If the language is unclear, one then looks to structure, purpose, and history to resolve that ambiguity.")

¹²¹ PHILIP BOBBITT, *CONSTITUTIONAL FATE: THEORY OF THE CONSTITUTION* 7, 123–24 (1982) (listing the traditional modalities as (1) historical, (2) textual, (3) structural, (4) prudential, and (5) doctrinal and arguing that a sixth, ethical modality, exists).

¹²² See Lawrence B. Solum, *The Interpretation-Construction Distinction*, 27 CONST. COMMENT. 95, 101–02 (2010).

requiring a normative theory which may include other modalities.¹²³ Of course, this process may result in conflict between results reached at the interpretation stage and those reached by reasoning via precedent (for example, where an important, related precedent is nonoriginalist), but it is precisely this issue that has attracted controversy as originalists debate how to fit nonoriginalist precedent into their system.¹²⁴ The rise of originalism is therefore further evidence of an intellectual trend among lawyers during the 1980s and '90s to make the law more rule-like and less subjective.

Regardless of the other ideological motives of certain proponents, originalism and textualism strike a similar, and certainly attractive, chord to multistep tests.¹²⁵ They add structure to inquiries that may seem unstructured to others, potentially adding to the predictability or at least the expressive rule-of-law values of the law.¹²⁶ Accordingly, the fact that stepification also emerged during this time suggests a connection between these trends in lawyerly thinking.

To restate the hypothesis, stepification emerged at this time to make the law more rule-like where a single rule would not do. That is, where the law could not be reduced to a single bright-line rule, it was rendered more procedural by instead replacing balancing tests with a series of *multiple* single inquiries. Stepification is, therefore, an effort to extend the reach of these new formalisms to more areas of the law. Indeed, in *City of Ladue v. Gilleo*, we see Justice Sandra Day O'Connor critiquing the majority for balancing rather than applying the tiers of scrutiny on the ground that the multistep tiers are themselves *rules* rather than "subjective balancing

¹²³ See *id.* at 104–07. This would include some combination of the doctrinal, prudential, and ethical modalities. See *id.*

¹²⁴ See, e.g., John O. McGinnis & Michael B. Rappaport, *Reconciling Originalism and Precedent*, 103 NW. U. L. REV. 803, 823–29 (2009) (developing a theory of how following precedent is consistent with originalism); Lawrence B. Solum, *Originalist Theory and Precedent: A Public Meaning Approach*, 33 CONST. COMMENT. 451, 456–70 (2018) (reviewing RANDY J. KOZEL, *SETTLED VERSUS RIGHT: A THEORY OF PRECEDENT* (2017)) (summarizing the problem of precedent for originalists and the debate on the topic and offering thoughts on how to resolve the debate).

¹²⁵ Textualism also arguably comes, in part, from the same law and economics sources I suggest were part of the drivers of stepification in general later. Eskridge & Frickey, *supra* note 114, at 340 ("The legal realists and legal process thinkers discredited intentionalism as a grand strategy for statutory interpretation; in its place they suggested purposivism. That theory has in turn been extensively criticized, especially by scholars influenced by the law and economics movement."); see *infra* Section II.B.2.

¹²⁶ Eskridge & Frickey, *supra* note 114, at 340 ("As suggested above, textualism appeals to the rule-of-law value that citizens ought to be able to read the statute books and know their rights and duties. By emphasizing the statutory words chosen by the legislature, rather than (what seem to be) more abstract and judicially malleable interpretive sources, textualism also appeals to the values of legislative supremacy and judicial restraint.").

tests.”¹²⁷ This theory has further resonance with Justice Antonin Scalia’s famous call for appellate courts to decide cases with rules rather than standards as crisp decision procedures were necessary to ensure uniformity in lower courts.¹²⁸ It is perhaps the case, then, that the language of multistep tests also reflects a shift to appellate courts seeing their role as more focused on formulating rules of decision rather than resolving individual disputes.¹²⁹ And where a bright-line rule would not do, stepification was the next best thing. For example, it may have been impossible to formulate a bright-line rule to govern deference to agency statutory interpretation, but at least the two-step *Chevron* formulation could add some rigidity to the inquiry.

Indeed, if the internal story is to be believed (and not cynically attributed to brute political considerations), perhaps the best parallel is to the story that Professor Gilmore tells about the move towards formalism during the height of tensions over slavery. As Professor Gilmore notes, a judge confronted with discomfort about the role she is charged to fill can either resign or choose to enforce the law in a way that limits the personal connection of the judge to the results she is articulating.¹³⁰ Of course, a retreat from the notion that judging is political is importantly different from the choice between resignation and enforcing law that one thinks is odious and unjust. While a principled judge may not resign in response to a belief that her job is becoming politicized, she may retreat to formalism as a refuge from the challenge that judging is personal, just as earlier judges did when charged with enforcing a regime that they felt was unjust.

Consequently, one ought not overread the degree to which formalism and sharp rules are associated with judicial neutrality. While the realist movement challenged the nature of judging, it is not obvious that a pre-stepified world was one in which judges were broadly perceived to be lawless, though it may look that way from certain current viewpoints.¹³¹ While it is difficult to unmoor ourselves from the modern day, it is not clear

¹²⁷ 512 U.S. 43, 59–60 (1994) (O’Connor, J., concurring) (arguing that despite the critiques of this rule-like approach, “it has substantial merit as well. It is a rule, in an area where fairly precise rules are better than more discretionary and more subjective balancing tests”). I am indebted to Professor Robert Post for suggesting this example.

¹²⁸ Antonin Scalia, *The Rule of Law as a Law of Rules*, 56 U. CHI. L. REV. 1175, 1178–79 (1989).

¹²⁹ I am indebted to Professor Michael Coenen for suggesting this connection.

¹³⁰ See GILMORE, *supra* note 109, at 33–34.

¹³¹ Indeed, lawyers had long tried to make law scientific before they discover the language of steps. See NEIL DUXBURY, *PATTERNS OF AMERICAN JURISPRUDENCE* 14 (1995) (“Langdell not only established the use of the case method as a pedagogical device, but also promoted the idea that the case method is necessary to the teaching of law as a science. It is the idea that law is a science, and the promotion of this idea by the use of the case method, which constitutes his sole yet fundamental contribution to American legal education.”); GILMORE, *supra* note 109, at 38–39 (observing how literally Langdell took this notion of law as a science).

that the development of stepification was a necessary development for the law to reform itself into a systematic endeavor. Indeed, Professor Gilmore documented the earlier movement in the United States towards formalism in light of the issues presented to judges in the shadow of slavery.¹³² But this earlier move, as observed above, was free of the stepification observed recently in American law.¹³³ The story thus cannot be one of a mere move towards formalism and away from the notion that judges imbued decisions with their personal, moral judgments.

Moreover, even to the extent that the internal story is correct, these conclusions are speculative and unfalsifiable. The causality story told here is not provable and, unlike arguments about the Supreme Court, cannot be verified with reference to the Justices' papers.¹³⁴ The connections suggest, however, that stepification resonates with broader trends in the legal community.

There is one further problem with this story. The problem with a simple story of lawyers moving to make the law more structured is that law has long been concerned with clarity and the development of workable rules.¹³⁵ Whether or not the language of steps improves clarity, an explanation that stands only on a story of clarity and discretion cannot be enough. There has long been a trend towards trying to make law more "scientific." This formalist move easily predates the move towards steps.¹³⁶ An internal story about a trend towards more rule-like law cannot tell us *why* the language of steps was chosen to fulfill this function. This answer must lie elsewhere.

2. *The External Story*

This Section makes the claim that stepification was not an isolated process within legal jurisprudence but one that is more fruitfully situated within a broader trend of proceduralization in the American elite. In particular, the rise of stepification is suggestively contemporaneous with the rise of decision-tree analysis within the field of management. The natural meeting place for these external trends and legal elites was the law and economics movement, which introduced a new attempt to render the law

¹³² GILMORE, *supra* note 109, at 32–35 (attributing the move towards formalism as part of the general desire of certain judges to distance themselves from the decisions they were required to issue in upholding the pro-slavery regime).

¹³³ See *supra* Section II.A.

¹³⁴ See, e.g., Cary Franklin, *The New Class Blindness*, 128 YALE L.J. 2, 32–35 (2018) (reviewing the papers of the Justices to argue that class was an underappreciated aspect of their decision-making with respect to certain rights decisions).

¹³⁵ Cf. GILMORE, *supra* note 109, at 23–32 (describing the movement in the early nineteenth century to codify the law in the name of uniformity).

¹³⁶ DUXBURY, *supra* note 131, at 24–25 (tracing the high-water mark of Langellian formalism to the *Restatement* movement in the early twentieth century).

scientific via the science of economics and its tools for deciding difficult normative issues.¹³⁷

Step tests are essentially decision-tree methods for resolving legal questions. Because they break up issues into component parts that can be resolved independently, these tests can be formulated and expressed as the sort of multipart decisions embodied in decision-tree analyses.¹³⁸

While decision-tree analysis may seem rather natural and therefore mundane today, it is in fact historically recent. In 1964, an article in the *Harvard Business Review* introduced the decision tree as a novel solution technique for management problems.¹³⁹ The article presents the decision tree as something of a managerial silver bullet, reducing complexity to a series of manageable individual decisions.¹⁴⁰ At once, the “choices, risks, objectives, monetary gains, and information needs involved in an investment problem” could be clarified and organized in a rigorous way.¹⁴¹ The author was careful to acknowledge that the decision tree could not eliminate entirely the need for judgment, but argued that by *structuring* the problem, the method could improve the quality of investments.¹⁴²

As Leigh Buchanan and Andrew O’Connell note, the term decision-making in business was borrowed from the field of public administration in the middle of the twentieth century.¹⁴³ Its introduction was not merely a stylistic change. Rather, Buchanan and O’Connell argue that these more formal methods altered management itself, introducing crispness and decisiveness into planning.¹⁴⁴

These methods were not without their critics. As quantitative methods infiltrated business school curricula, commentators charged that they had

¹³⁷ See *infra* notes 157–162 and accompanying text.

¹³⁸ See Re, *supra* note 59, at 611 fig.1 (expressing *Chevron* in a decision-tree format).

¹³⁹ See John F. Magee, *Decision Trees for Decision Making*, HARV. BUS. REV., July 1964, at 126, 127 (“In this article I shall present one recently developed concept called the ‘decision tree,’ which has tremendous potential as a decision-making tool. The decision tree can clarify for management, as can no other analytical tool that I know of, the choices, risks, objectives, monetary gains, and information needs involved in an investment problem. We shall be hearing a great deal about decision trees in the years ahead. Although a novelty to most businessmen today, they will surely be in common management parlance before many more years have passed.”).

¹⁴⁰ See *id.* at 127–28.

¹⁴¹ *Id.* at 127.

¹⁴² See *id.* at 135.

¹⁴³ Leigh Buchanan & Andrew O’Connell, *A Brief History of Decision Making*, HARV. BUS. REV., Jan. 2006, at 32, 33.

¹⁴⁴ *Id.*

displaced the “practice” of business.¹⁴⁵ This critique interestingly parallels the argument against certain manifestations of formalism. Critics of the new scientific methods argued that they were too abstract, prizing a level of “analytic detachment and methodological elegance over insight based on experience.”¹⁴⁶

The debate over scientific management then interestingly parallels legal debates over more rigid decision-making methods. Proponents of rigidity argue that it disciplines decisions.¹⁴⁷ But critics argue it strips out the valuable role of individual judgment.¹⁴⁸ Despite these concerns, however, it seems that for now the tools of technical management are mostly here to stay.¹⁴⁹

The interest in using decision theory to improve judgment was not limited to business. For example, in 1951 Professor John Rawls published a paper aiming to develop a seven-step decision theory for ethical judgments.¹⁵⁰ Professor Rawls attempted this work because, in his words:

[T]he objectivity or the subjectivity of moral knowledge turns, not on the question whether ideal value entities exist or whether moral judgments are caused by emotions or whether there is a variety of moral codes the world over, but simply on the question: does there exist a reasonable method for validating and invalidating given or proposed moral rules and those decisions made on the basis of them?¹⁵¹

For morality to be objective in Professor Rawls’s telling, it must be so in the sense that science is objective, with rules to decide between true and false claims.¹⁵²

¹⁴⁵ Samuel Paul, *Management Education: Emerging Trends*, VIKALPA, Oct.–Dec. 1992, at 11, 14 (“By the 1980s, the pendulum had swung the other way and some schools which were using sophisticated quantitative techniques were being attacked for putting too much emphasis on theoretical knowledge about business and for seriously neglecting the ‘practice’ of business.”); see David J. Teece & Sidney G. Winter, *The Limits of Neoclassical Theory in Management Education*, 74 AM. ECON. REV. 116, 116 (1984) (discussing this critique).

¹⁴⁶ Robert H. Hayes & William J. Abernathy, *Managing Our Way to Economic Decline*, HARV. BUS. REV., July–Aug. 1980, at 67, 70.

¹⁴⁷ See, e.g., Magee, *supra* note 139, at 135 (“Using the decision tree, management can consider various courses of action with greater ease and clarity. The interactions between present decision alternatives, uncertain events, and future choices and their results become more visible.”).

¹⁴⁸ See, e.g., Hayes & Abernathy, *supra* note 146, at 70; Paul, *supra* note 145, at 14.

¹⁴⁹ See, e.g., Jay Peters, *Whole Foods Is Reportedly Using a Heat Map to Track Stores at Risk of Unionization*, VERGE (Apr. 20, 2020, 2:50 PM), <https://www.theverge.com/2020/4/20/21228324/amazon-whole-foods-unionization-heat-map-union> [<https://perma.cc/W78Q-7QDA>] (reporting on how Whole Foods is using a heat map, which uses more than two dozen different metrics, to track stores at risk of unionization).

¹⁵⁰ John Rawls, *Outline of a Decision Procedure for Ethics*, 60 PHIL. REV. 177, 191–93 (1951).

¹⁵¹ *Id.* at 177.

¹⁵² *Id.*

At the same time, there was a parallel move in poetry criticism towards a new formalism and away from more open forms.¹⁵³ Writing in 1987, Professor Dana Gioia observed that in the 1960s it “was a truth universally acknowledged that a young poet in possession of a good ear would want to write free verse.”¹⁵⁴ The revival of form at the tail end of the ’70s came after “knowing critics had declared rhyme and meter permanently defunct.”¹⁵⁵ Of course, in the case of poetry, this move was in part a return to traditional approaches and not the novel innovation I am arguing stepification represented. But there is an important similarity in the appreciation of more closed and regimented forms, one suggestive of the intellectual currents of the time.¹⁵⁶

Relevant as well is the movement towards law and economics with its focus on deriving objective standards for law.¹⁵⁷ Perhaps the most appropriate summation of the goals of that movement was articulated by then-Professor Richard Posner in the inaugural issue of the *Journal of Legal Studies*:

The aim of the Journal is to encourage the application of scientific methods to the study of the legal system. As biology is to living organisms, astronomy to the stars, or economics to the price system, so should legal studies be to the legal system: an endeavor to make precise, objective, and systematic observations of how the legal system operates in fact and to discover and explain the recurrent patterns in the observations—the “laws” of the system.¹⁵⁸

In the vision of law and economics, the law was to return to a scientific state, one that was precise and objective.¹⁵⁹

Indeed, the first issue of the *Journal* specifically referenced the need for a theory of legal decision-making as one of the core themes of its mission.¹⁶⁰ Professor Posner wrote that “[m]any of the practices and institutions by

¹⁵³ I am indebted to Nicholas Parrillo for pointing out the concurrent rise of formalism in literature and literary theory.

¹⁵⁴ Dana Gioia, *Notes on the New Formalism*, 40 HUDSON REV. 395, 395 (1987).

¹⁵⁵ *Id.*

¹⁵⁶ Cf. Alan Shapiro, *The New Formalism*, 14 CRITICAL INQUIRY 200, 211–12 (1987) (discussing the view that formalism in poetry was inherently conservative while open forms were viewed as more “open, organic, exploratory, [and] natural”).

¹⁵⁷ See Philip Bobbitt, *The Age of Consent*, in GILMORE, *supra* note 109, at 100, 112–13.

¹⁵⁸ Richard A. Posner, *Volume One of The Journal of Legal Studies—an Afterword*, 1 J. LEGAL STUD. 437, 437 (1972).

¹⁵⁹ It was out of this general trend that a game theoretic view of the law could even emerge. See DOUGLAS G. BAIRD, ROBERT H. GERTNER & RANDAL C. PICKER, *GAME THEORY AND THE LAW* 1–2 (1994) (arguing that “a rigorous focus on strategic behavior” can advance an understanding of the law in contexts ranging from simple tort dilemmas to complex antitrust suits).

¹⁶⁰ Posner, *supra* note 158, at 439.

which legal rules are formulated and applied are poorly understood,” which the articles in the first volume of the new *Journal* aimed to remedy.¹⁶¹ While none of these articles addressed the form of legal tests head on, Professor Posner declared the study of legal rules and tests as one of the three missions of his initial volume.¹⁶²

Seen against this backdrop, the history of stepification seems less surprising and more of an inevitable evolution. Stepwise tests carry with them an air of formulaic decision-making, the precise element that the law and economics movement was looking for. Moreover, the addition of steps and the language of decision theory to law was in line with a broader elite movement towards structure and away from amorphous judgment.¹⁶³ The formalization of judgment into mathematical rules occurred in multiple areas at similar times, impacting each.

While the above examples are suggestive of the underlying drivers of stepification, they cannot be dispositive. It is impossible to identify precisely why judges and lawyers chose to take up the language of decision trees and multistep inquiries to guide their work. The most that can be said is that the tool they turned to was one that had currency in other areas as a tool to improve judgment by adding procedural structure.

Thus, while elements of the external story for the rise of stepification seem somewhat convincing,¹⁶⁴ that story is critically incomplete. This Article is not about the effort to make the law more neutral or about that project’s success or failure. Rather, what is important is the form that effort took. In adopting stepification, lawyers drew on a broader movement among elites of the time to rationalize decision-making more generally and to reduce the ambiguity of decisions to workable decision procedures. The question then becomes the implication of this historical turn.

¹⁶¹ *Id.* at 439–40.

¹⁶² *Id.* (“A third important theme of this volume is the quest for a theory of legal decision-making. Many of the practices and institutions by which legal rules are formulated and applied are poorly understood—the jury, the rhetoric of the appellate opinion, the formalities of the adversary process, are some important examples. A number of articles in volume one explore perplexing features of the legal decisional process.”).

¹⁶³ Though, perhaps ironically, the business community eventually experienced a shift back towards the cult of the “gut.” See Buchanan & O’Connell, *supra* note 143, at 40 (“That semantic shift—from human’s stomach to lion’s heart—helps explain the current fascination with gut decision making. From our admiration for entrepreneurs and firefighters, to the popularity of books by Malcolm Gladwell and Gary Klein, to the outcomes of the last two U.S. presidential elections, instinct appears ascendant. Pragmatists act on evidence. Heroes act on guts.”). Still, the practice of modern management shows that the idea of more precise and scientific decision-making influences current practices yet.

¹⁶⁴ And certainly there was a real drive by certain actors to make the law more neutral by making it more scientific. Bobbitt, *supra* note 157, at 110–13 (describing the economic turn as an effort to remove politics from law).

* * *

Stepification as a process is not an isolated feature of individual areas of the law. Rather, it is ubiquitous across different areas of the law. Where chaos appears within the law, stepification serves to organize it. This organization is sometimes illusory, however, and chaos eventually returns as pressure is put upon various doctrinal frameworks. Step tests can break down as their normative tensions expose certain weaknesses in their ability to capture what is important about the underlying issue in question. To see this, the next Part of this Article works through some concrete examples of the phenomenon.

III. EXAMPLES OF STEPIFICATION (AND ITS DRAWBACKS)

While the above overview of stepification demonstrates its increasing role in the law, its normative implications are difficult to observe without concrete examples. As such, this Part discusses four brief case studies of the process in action—two in the past and two more recent examples—to illustrate the claim that stepification has universal qualities and significance across areas of the law.

The first two examples are historical regimes: *Chevron* and *Lemon*. In both cases, the Supreme Court articulated what seemed to be a universal framework for the issue at hand. And in both cases, the structure has persisted despite challenges to its legitimacy.

The final two examples are cases in which the Court has chosen not to adopt a multistep framework. I first include the example of criminal procedure because it is in these cases that the Court has most openly grappled with the implications of adopting multistep tests. Additionally, in partisan gerrymandering cases, the majority declined to adopt a multistep framework despite the argument by the dissenters that a multistep approach was both workable and familiar to courts.

The purpose of these examples is to show stepification is transdoctrinal and transhistorical. Of course, other options could have been chosen, such as the qualified immunity two-step test, but these examples illustrate important elements of what I argue is a general phenomenon.

A. Chevron

The most famous example of stepification is the transition from the deference regime announced in *Skidmore v. Swift & Co.*, in which deference was based on the agency's "power to persuade,"¹⁶⁵ to the two-step test of

¹⁶⁵ 323 U.S. 134, 140 (1944).

Chevron.¹⁶⁶ *Skidmore* provided a general set of factors that a court would consider, stating that deference would depend on “the thoroughness evident in its consideration, the validity of its reasoning, its consistency with earlier and later pronouncements, and all those factors which give it power to persuade, if lacking power to control.”¹⁶⁷ In contrast, the *Chevron* doctrine asks (1) whether the statutory term is ambiguous and then (2) whether the agency’s interpretation of the ambiguity is reasonable.¹⁶⁸ While dispute exists about whether *Chevron* actually matters at the Supreme Court level, it has been treated as a revolution in the area of agency interpretation.¹⁶⁹

Whatever the precise impact of the decision, *Chevron* has marked an important change in the form of the deference inquiry. The *Skidmore* test was a case-by-case balancing approach in which deference was based on a variety of judgments about the weight of the agency’s interpretation.¹⁷⁰ In contrast, *Chevron* purports to be much simpler, condensing this analysis into two relatively clean questions.¹⁷¹ Moreover, the questions are based on analysis of the statute, rather than fraught balancing judgments about the persuasiveness of the agency interpretation.

What is interesting is that *Chevron* has not remained two steps in practice. The complex (and normatively fraught) issues in administrative law have resulted in subsequent modifications to the doctrine. For example, in place of *Skidmore*’s evaluation of the thoroughness of the agency’s consideration of a given interpretation,¹⁷² the Court has created “Step Zero”¹⁷³

¹⁶⁶ Professor Cary Coglianese, however, has argued that *Chevron* actually contains a number of “interstitial steps” that structure the space between Step One and Step Two. See Cary Coglianese, *Chevron’s Interstitial Steps*, 85 GEO. WASH. L. REV. 1339, 1339–40 (2017).

¹⁶⁷ 323 U.S. at 140.

¹⁶⁸ *Chevron U.S.A. Inc. v. Nat. Res. Def. Council, Inc.*, 467 U.S. 837, 842–44 (1984).

¹⁶⁹ William N. Eskridge, Jr. & Lauren E. Baer, *The Continuum of Deference: Supreme Court Treatment of Agency Statutory Interpretations from Chevron to Hamdan*, 96 GEO. L.J. 1083, 1120–22 (2008) (contending that *Chevron* has had significantly less impact than its status would suggest); Orin S. Kerr, *Shedding Light on Chevron: An Empirical Study of the Chevron Doctrine in the U.S. Courts of Appeals*, 15 YALE J. ON REGUL. 1, 30–31 (1998) (finding that *Chevron* resulted in considerable deference to agency interpretations).

¹⁷⁰ *Skidmore*, 323 U.S. at 140.

¹⁷¹ Though, famously, some have argued that *Chevron* has only one step. Stephenson & Vermeule, *supra* note 55, at 598–600.

¹⁷² *Skidmore*, 323 U.S. at 140.

¹⁷³ See Merrill & Hickman, *supra* note 41, at 873.

through its decisions in *Mead*¹⁷⁴ and *Barnhart*.¹⁷⁵ The new Step Zero is a threshold inquiry into whether the agency action is even entitled to claim the benefits of *Chevron* deference at all, a new step that the court must ask before applying the usual two steps. In creating such a step, the Court has reinjected the consideration of the nature of the executive decision into the inquiry, seemingly recognizing the need to restore some aspects of the old *Skidmore* regime.

Moreover, through the creation of the major questions doctrine, which functions as an additional off-ramp from *Chevron* (and thus as arguably another implicit “step”), the Court has similarly recognized an additional shortcoming of the stepified test.¹⁷⁶ It has fixed this shortcoming, however, by arguably adding a further step to the test. In *Brown & Williamson*, the Court seemed to recognize that there are situations in which an agency interpretation passes *Chevron* but in which courts should still not defer to that interpretation.¹⁷⁷ This can be read as an admission that despite the stepified simplicity of the two-step *Chevron* form, its literal application can occasionally oversimplify hard cases that must then be addressed through reference to the normative values that the step test imperfectly captures.¹⁷⁸ The major questions doctrine then functions as an additional step to the inquiry, whereby the court can choose not to apply *Chevron* (or to disregard its result) if another condition is met.

¹⁷⁴ *United States v. Mead Corp.*, 533 U.S. 218, 226–27 (2001) (holding that in order for an agency interpretation to qualify for *Chevron* deference, Congress must have “delegated authority to the agency generally to make rules carrying the force of law” and the agency interpretation must have been “promulgated in the exercise of that authority”).

¹⁷⁵ *Barnhart v. Walton*, 535 U.S. 212, 222 (2002) (“In this case, the interstitial nature of the legal question, the related expertise of the Agency, the importance of the question to administration of the statute, the complexity of that administration, and the careful consideration the Agency has given the question over a long period of time all indicate that *Chevron* provides the appropriate legal lens through which to view the legality of the Agency interpretation here at issue.”).

¹⁷⁶ *FDA v. Brown & Williamson Tobacco Corp.*, 529 U.S. 120, 159–61 (2000) (articulating the major questions doctrine); see also Russell L. Weaver, *Some Realism About Chevron*, 58 MO. L. REV. 129, 167–74 (1993) (cataloging ways in which the Court had avoided applying *Chevron*).

¹⁷⁷ To be sure, there is some dispute about the precise birthdate of the major questions exception. See Abigail R. Moncrieff, *Reincarnating the “Major Questions” Exception to Chevron Deference as a Doctrine of Noninterference (or Why Massachusetts v. EPA Got It Wrong)*, 60 ADMIN. L. REV. 593, 598–606 (2008) (providing an overview of the cases). Professors Michael Coenen and Seth Davis have argued, in particular, that *Brown & Williamson*, 529 U.S. at 120, is really a Step One case in which majorness is used as an interpretive canon. Michael Coenen & Seth Davis, *Minor Courts, Major Questions*, 70 VAND. L. REV. 777, 787–88 (2017). Their narrative suggests that it was really *King v. Burwell*, 576 U.S. 473 (2015), that recognized that majorness might give rise to questions in which *Chevron* might simply not apply. See Coenen & Davis, *supra*, at 793.

¹⁷⁸ See *supra* Part II. In this context, the typical hard case I have in mind is one in which the literal application of the doctrine may sit uneasily with other legal considerations. In contrast, an easy case is one in which the straightforward application of the doctrine leads to the “right” result.

Indeed, the creation of these exceptions to *Chevron* has still not been enough for some commentators. For example, Justice Brett Kavanaugh has proposed adding additional structure to the first step of *Chevron* to discipline its analysis.¹⁷⁹ In response to *Chevron*'s encouragement of aggressive policymaking,¹⁸⁰ Justice Kavanaugh has suggested reorienting *Chevron* around an inquiry into whether the statutory language in question has a "best reading."¹⁸¹ If so, Justice Kavanaugh suggests that the Court should follow this reading.¹⁸² Only in cases of "statutes using broad and open-ended terms like 'reasonable,' 'appropriate,' 'feasible,' or 'practicable'" should judges defer to the policy choice of the agency.¹⁸³ Essentially, Justice Kavanaugh's proposal would alter the shape of Step One, replacing it with a new inquiry that would be centered around whether the text is open-ended.

It is not just that *Chevron* has encountered cases that push at its limits, but that its changes have taken the form of new steps in the analysis.¹⁸⁴ For now, the Court has not retreated to a balancing approach but has instead attempted to reinforce the multistep analysis via further stepification. Whether this trend will continue is unclear, but it is telling that the Court continues to proceed in this way in response to challenges.

Of course, the form of *Chevron* has been remarked on in the past: Professor Re has argued that it has important normative implications because the different statements of *Chevron* as either a one- or two-step test affect the normative implications it conveys.¹⁸⁵ Similarly, Professor Cary Coglianese has argued that the two-step formulation has important implications for interbranch dialogue as the two distinct steps communicate information to agencies about the scope of their authority.¹⁸⁶ My point in this Article is not to disagree with Professor Re's assessment that the form of the stepified test has important implications—it certainly does—but rather to question the stepified form more generally. As Professor Re's article admits,

¹⁷⁹ Kavanaugh, *supra* note 11, at 2150–54.

¹⁸⁰ *Id.* at 2150–51.

¹⁸¹ *Id.* at 2154.

¹⁸² *Id.*

¹⁸³ *Id.* at 2153–54.

¹⁸⁴ A depiction of these shifts is shown in *infra* Figure 5.

¹⁸⁵ Re, *supra* note 59, at 608 ("On reflection, there are important advantages and disadvantages to traditional *Chevron*'s command that courts should ask about both mandatoriness and reasonableness in every case. For example, requiring courts to answer both questions facilitates the rapid development of the law, but asking only about reasonableness seems consistent with principles of judicial restraint. Instead of following traditional two-step, perhaps courts should ask only about reasonableness.").

¹⁸⁶ Coglianese, *supra* note 166, at 1374–86. This argument resembles in interesting ways the argument I make below about the communicative benefits of multistep tests for appellate courts looking to provide guidance to lower courts. See *infra* Section IV.A.

the question of what steps are chosen does matter to the law. The subsequent examples ask whether there should be steps at all.

B. Lemon

A useful alternative example is the Supreme Court's *Lemon* framework for evaluating Establishment Clause challenges.¹⁸⁷ In *Lemon*, the Court held that state action is consistent with the Establishment Clause when it meets three prongs. First, the action in question must have a secular purpose.¹⁸⁸ Second, its "principal or primary effect must be one that neither advances nor inhibits religion."¹⁸⁹ Third, the action in question "must not foster 'an excessive government entanglement with religion.'"¹⁹⁰

While this framework seems to present a workable method to address these claims, it has been under significant pressure. This is perhaps because while the steps as articulated seem to address independent issues, they may not be as separate as they seem.¹⁹¹ In particular, difficult Establishment Clause issues involve combinations of these considerations, suggesting that the underlying normative issue—when does the state impermissibly impose religious belief on individuals—cannot be neatly divided up into three truly independent elements.

One example of this pressure is the application of the test to passive monuments. In particular, in *Van Orden v. Perry*, the Justices encountered the case of a long-standing monument in the shape of the Ten Commandments on the Texas State Capitol grounds.¹⁹² However, rather than apply *Lemon*, four Justices would have held that *Lemon* is inapplicable to monuments such as the one in question.¹⁹³ The plurality stopped short of overruling *Lemon* entirely (a move that would have been strange given that another case decided the same day applied the test¹⁹⁴), but the plurality significantly limited its effect. The controlling opinion in *Van Orden*, Justice

¹⁸⁷ *Lemon v. Kurtzman*, 403 U.S. 602, 612–13 (1971).

¹⁸⁸ *Id.* at 612.

¹⁸⁹ *Id.* (citing *Bd. of Educ. of Cent. Sch. Dist. No. 1 v. Allen*, 392 U.S. 236, 243 (1968)).

¹⁹⁰ *Id.* at 613 (quoting *Walz v. Tax Comm'n*, 397 U.S. 664, 674 (1970)).

¹⁹¹ See *infra* Section IV.B.

¹⁹² 545 U.S. 677, 681 (2005) (plurality opinion).

¹⁹³ *Id.* at 686 ("Whatever may be the fate of the *Lemon* test in the larger scheme of Establishment Clause jurisprudence, we think it not useful in dealing with the sort of passive monument that Texas has erected on its Capitol grounds. Instead, our analysis is driven both by the nature of the monument and by our Nation's history.").

¹⁹⁴ *McCreary County v. ACLU of Ky.*, 545 U.S. 844, 859–66 (2005) (declining to limit *Lemon*).

Stephen Breyer's concurrence, looked to *Lemon* only as a guidepost, not as the correct doctrinal test.¹⁹⁵

Indeed, in a recent Term, the Supreme Court was asked once again to reconsider *Lemon*'s application to a passive monument, this time in the form of a thirty-foot cross in *American Humanist Association*.¹⁹⁶ This case again seemed to present the hard question of what to do when a rigid application of *Lemon* would require the Court to hold that a nearly 100-year-old monument must be torn down.¹⁹⁷ In considering the case, four Justices highlighted the test's fraught history, noting that "[i]n many cases, this Court has either expressly declined to apply the test or has simply ignored it" despite the *Lemon* Court's ambition to "provide a framework for all future Establishment Clause decisions."¹⁹⁸ Those Justices continued to critique the test's "shortcomings" but there was no fifth vote to discard the test.¹⁹⁹ Still, the Court ultimately decided that the cross did not violate the Establishment Clause, further weakening *Lemon*.

Lemon's history highlights both the ambitions and drawbacks of multistep tests, particularly at the Supreme Court level. Its creators attempted to craft a framework for all Establishment Clause challenges that could be easily applied. While *Lemon* captures certain essentials about the Establishment Clause, a rigid application of its steps leads occasionally to results that a majority of the Court considers unacceptable. As a result, while there is no majority to discard the test, the Court is willing to disregard it in particular cases where its straightforward application would lead to the "wrong" result.

This demonstrates the limitations of step tests, especially when the Court is quick to ignore them in certain circumstances. *Lemon* might be seen as a case in which the impulse to embrace a multistep approach has led to a test the Supreme Court feels comfortable regularly departing from, even as a majority of the Court has been reluctant to depart from *Lemon* once and for all. It may be that *Lemon* serves other goals for the Court, perhaps by

¹⁹⁵ 545 U.S. at 700 (Breyer, J., concurring in the judgment) (characterizing *Lemon* as a useful "guidepost[]" while arguing that "no exact formula can dictate a resolution to such fact-intensive cases").

¹⁹⁶ See Brief for Petitioner Maryland-National Capital Park & Planning Commission at 22, *Am. Legion v. Am. Humanist Ass'n*, 139 S. Ct. 2067 (2019) (No. 18-18) ("Because the basic principles of the Court's Establishment Clause precedents lead to a clear result, the Court need not apply the *Lemon* test here.").

¹⁹⁷ 139 S. Ct. at 2074.

¹⁹⁸ *Id.* at 2080 (plurality opinion).

¹⁹⁹ *Id.* at 2080–82; see also *id.* at 2094 (Kagan, J., concurring in part) ("Although I agree that rigid application of the *Lemon* test does not solve every Establishment Clause problem, I think that test's focus on purposes and effects is crucial in evaluating government action in this sphere—as this very suit shows. I therefore do not join Part II–A.").

providing useful guideposts for the lower courts. But its history at the Supreme Court suggests that creation of a multistep regime to govern highly contested and complicated cases can be a fraught endeavor.

C. *Constitutional Criminal Procedure*

Interestingly, constitutional criminal procedure has featured open discussion of the merits of multistep tests and their value to the law. In this space, the Court has openly addressed the utility of these tests and whether they appropriately fit the underlying needs of the law. In particular, two examples in this space illustrate the dynamics of stepification: the evolution of the *Katz* test and the rejection of the two-prong test for whether informants' tips constitute probable cause.

Of these examples, the test of *Katz v. United States*²⁰⁰ is undoubtedly more famous. In *Katz*, the Court attempted to address how the Fourth Amendment applied to government surveillance of the defendant's conversation in a public telephone booth.²⁰¹ Because the telephone booth was outside of the home and arguably in public, the question was how the Fourth Amendment would apply to this new space. While the Court ultimately held that the call was protected,²⁰² the case, however, is more famous for Justice John Marshall Harlan's concurrence, which stated what would eventually become an important test for determining the scope of the Fourth Amendment. Harlan's formulation imposed a "twofold requirement" for the Amendment's protections to apply: "[F]irst that a person have exhibited an actual (subjective) expectation of privacy and, second, that the expectation be one that society is prepared to recognize as 'reasonable.'"²⁰³

While *Katz* continues to be cited approvingly by the Court, its subsequent history reveals two interesting facets of multistep tests.²⁰⁴ The first is that its use is subject to qualification because the *Katz* test is an imperfect fit for how the Court conceptualizes the Fourth Amendment. For example, the Court has indicated that the *Katz* test is not coextensive with the coverage of the Fourth Amendment and instead serves to supplement the

²⁰⁰ 389 U.S. 347 (1967).

²⁰¹ *Id.* at 348–49.

²⁰² *Id.* at 359 (“These considerations do not vanish when the search in question is transferred from the setting of a home, an office, or a hotel room to that of a telephone booth. Wherever a man may be, he is entitled to know that he will remain free from unreasonable searches and seizures.”).

²⁰³ *Id.* at 361 (Harlan, J., concurring).

²⁰⁴ *See, e.g.,* *Carpenter v. United States*, 138 S. Ct. 2206, 2213 (2018) (citing *Katz* as articulating an important dimension of the Fourth Amendment's protections).

Amendment's protection of property.²⁰⁵ In particular, while the *Katz* test describes a set of interests protected by the Fourth Amendment, the Amendment's protections stretch beyond *Katz* to cover other categories of intrusions. Thus, in effect, an additional step has been added to the test. Arguably, one must first look to whether some prior interest covers the case. If the answer to the first question is no, then the court will proceed to apply the *Katz* framework. While not fatal, the need to add this additional qualification indicates the difficulty courts face when attempting to formulate comprehensive frameworks in this way.²⁰⁶

The second, and more important, issue is that over time, the analysis under such a test can become wooden. Because multistep tests allow each factor to be viewed in isolation, the analysis can become detached from the underlying value the test is supposed to capture.²⁰⁷ This risk is not academic; in fact, Justice Harlan himself observed it. In *United States v. White*, Justice Harlan, in discussing *Katz*, admonished the Court that “[w]hile . . . formulations [like *Katz*] represent an advance over the unsophisticated trespass analysis of the common law, they too have their limitations and can, ultimately, lead to the substitution of words for analysis.”²⁰⁸ These frameworks have the potential to make decisions more structured, but they do so at the risk of potentially making decisions *too* easy for judges who can march through the steps of a test without actually engaging with the normative issues in the case.²⁰⁹

The issue of rigidity is addressed in a second issue in criminal law: how to treat the reliability of informants' tips. In particular, when are the tips of informants sufficient to create probable cause for police action? In *Illinois v. Gates*, the Supreme Court confronted the application of a “two-prong test”

²⁰⁵ See *Florida v. Jardines*, 569 U.S. 1, 5 (2013) (“[T]hough *Katz* may add to the baseline, it does not subtract anything from the Amendment’s protections ‘when the Government *does* engage in [a] physical intrusion of a constitutionally protected area’” (quoting *United States v. Knotts*, 460 U.S. 276, 286 (1983) (Brennan, J., concurring in the judgment))). This is leaving aside those who argue that the test should be discarded entirely. *Carpenter*, 138 S. Ct. at 2236 (Thomas, J., dissenting) (“The *Katz* test has no basis in the text or history of the Fourth Amendment. And, it invites courts to make judgments about policy, not law. Until we confront the problems with this test, *Katz* will continue to distort Fourth Amendment jurisprudence.”); *id.* at 2261–72 (Gorsuch, J., dissenting) (“I would look to a more traditional Fourth Amendment approach. Even if *Katz* may still supply one way to prove a Fourth Amendment interest, it has never been the only way. Neglecting more traditional approaches may mean failing to vindicate the full protections of the Fourth Amendment.”).

²⁰⁶ See *infra* Part V (discussing the need to increase the complexity of tests over time to attempt to fit them appropriately to the values the doctrine aims to implement).

²⁰⁷ See *infra* Section IV.B.

²⁰⁸ 401 U.S. 745, 786 (1971) (Harlan, J., dissenting).

²⁰⁹ See *infra* Part IV.

that had been developed by courts in light of the Supreme Court's earlier decisions addressing this issue.²¹⁰

Instead of embracing the two-prong test, the Court rejected it in favor of a holistic "totality-of-the-circumstances" approach.²¹¹ The Court held that the nature of probable cause as "a fluid concept—turning on the assessment of probabilities in particular factual contexts" rendered it "not readily, or even usefully, reduced to a neat set of legal rules."²¹²

Moreover, the Court critiqued how the test separated the two issues of "reliability" and "basis of knowledge" into "independent channels," stating:

Instead, they are better understood as relevant considerations in the totality-of-the-circumstances analysis that traditionally has guided probable-cause determinations: a deficiency in one may be compensated for, in determining the overall reliability of a tip, by a strong showing as to the other, or by some other indicia of reliability.²¹³

The Court noted that a strong showing on one factor could plausibly lead to the conclusion that only a weaker showing on the other was necessary.²¹⁴ In other words, the separation of the individual factors could not be justified with reference to the underlying normative issue.²¹⁵ Therefore, the Court decided to abandon the test in favor of the holistic approach to the issue in which all factors were examined together.

Both of these discussions by the Court are exceptionally frank about the function of multistep tools as methods of jurisprudence. Indeed, these opinions largely foreshadow the later discussion of the potential pitfalls inherent in these tests.²¹⁶ Moreover, while the *Katz* framework continues to be deployed, the rejection of the two-step analysis in *Gates* illustrates how these frameworks can collapse as well.²¹⁷

²¹⁰ 462 U.S. 213, 228 (1983) ("The Illinois Supreme Court, like some others, apparently understood *Spinelli* as requiring that the [tip] satisfy each of two independent requirements before it could be relied on."). This test required that the tip "first had to adequately reveal the 'basis of knowledge' of the [informant]—the particular means by which he came by the information given in his report. Second, it had to provide facts sufficiently establishing either the 'veracity' of the affiant's informant, or, alternatively, the 'reliability' of the informant's report in this particular case." *Id.* at 228–29.

²¹¹ *Id.* at 230–31.

²¹² *Id.* at 232.

²¹³ *Id.* at 233.

²¹⁴ *Id.* at 233–34.

²¹⁵ See *infra* Section IV.B.

²¹⁶ Though the initial decisions of courts to adopt the two-prong approach later rejected by *Gates* arguably shows the appeal of these methods. See *infra* Section IV.A (discussing the benefits of stepified tests).

²¹⁷ See *infra* Section IV.B.

D. Partisan Gerrymandering

While the above examples capture the challenges that stepification creates when the new multistep tests are unable to effectively express the normative issues at stake, they also obscure some of the rhetorical advantages that stepification offers. This dynamic can be observed in the deployment of stepification as a strategy—though ultimately unsuccessful at the Supreme Court—to convince courts that workable standards exist to address partisan gerrymandering.²¹⁸

The arc of efforts to reform partisan gerrymandering has largely been one of asking whether judicially manageable standards exist to resolve these cases.²¹⁹ In *Vieth*, four Justices concluded that no such standards existed,²²⁰ but Justice Anthony Kennedy consciously left the door open to the possibility that such standards might emerge at some later time.²²¹ Consequently, advocates of judicial intervention in partisan gerrymandering searched for standards that would meet the bar of manageability and, in two cases decided last Term, presented several possibilities to the Court that they claimed were manageable.²²²

In particular, plaintiffs advanced two similar three-prong tests in the lower courts to deal with partisan gerrymanders.²²³ Both tests essentially asked courts to find a gerrymander unconstitutional if it is (1) motivated by intent to discriminate against individuals based on their past voting behavior, (2) has the effect of actually discriminating against these voters, and

²¹⁸ See *Rucho v. Common Cause*, 139 S. Ct. 2484, 2502, 2506–07 (2019) (deciding that partisan gerrymandering cases are nonjusticiable and rejecting the tests proposed to address the issue as insufficiently “judicially discernible and manageable”).

²¹⁹ Specifically, the *Vieth* plurality argued that one limitation placed on courts is that “law pronounced by the courts must be principled, rational, and based upon reasoned distinctions.” *Vieth v. Jubelirer*, 541 U.S. 267, 278 (2004) (plurality opinion); see *Baker v. Carr*, 369 U.S. 186, 217 (1962) (“Prominent on the surface of any case held to involve a political question is found a textually demonstrable constitutional commitment of the issue to a coordinate political department; or a lack of judicially discoverable and manageable standards for resolving it . . .”).

²²⁰ 541 U.S. at 281 (plurality opinion).

²²¹ *Id.* at 306 (Kennedy, J., concurring in the judgment) (“[W]hile understanding that great caution is necessary when approaching this subject, I would not foreclose all possibility of judicial relief if some limited and precise rationale were found to correct an established violation of the Constitution in some redistricting cases.”).

²²² See, e.g., Brief for Appellees League of Women Voters of North Carolina at 50–63, *Rucho*, 139 S. Ct. 2484 (No. 18-422) (arguing that the three-part framework of intent, effect and justification meets the standard for manageability); Brief for Appellees at 31–42, *Lamone v. Benisek*, No. 18-726 (U.S. Mar. 4, 2019) (same).

²²³ Though note that arguably the tests in question are the same. *Rucho*, 139 S. Ct. at 2516 (Kagan, J., dissenting) (“And both courts (like others around the country) used basically the same three-part test to decide whether the plaintiffs had made out a vote dilution claim. As many legal standards do, that test has three parts: (1) intent; (2) effects; and (3) causation.”).

(3) cannot be justified based on the state's geography or some other legitimate objective.²²⁴ This test found some success as a workable standard in the lower courts. District courts used it to find partisan gerrymandering cases justiciable and to rule against extreme cases of gerrymandering, thinking that a test of this form might prove to be the workable standard for which the Supreme Court was searching.²²⁵

In *Common Cause v. Rucho*, the League of Women Voters focused on this three-prong test, arguing that its structure creates a workable framework for courts to determine which gerrymanders violate the Constitution.²²⁶ Similarly, the appellees in *Benisek v. Lamone* argued that this test represents “an appropriate fit” for the court to evaluate these claims.²²⁷

Examining the briefs in the case shows the work that the rhetoric of stepification is doing in these cases. For example, the appellees in the *Rucho* case argued that each prong allows the Court to limit the number of gerrymanders that would be potentially unconstitutional. In their telling, the intent prong served to restrict the ability of plaintiffs to challenge maps that incidentally confer an advantage to one side,²²⁸ the effects prong ensured that districts must be considered within the context of the entire map,²²⁹ and the justification prong allowed courts to deny challenges if a set of randomly generated maps show that the map in question is within the norm for the state.²³⁰

Here, the use of the step test allowed the appellees to show that the need to prove each and every prong of their test makes the standard itself manageable. This formulation of the test, however, comes at the cost of not directly addressing the root of the issue. A map that has significant

²²⁴ See *Common Cause v. Rucho*, 318 F. Supp. 3d 777, 861–68 (M.D.N.C. 2018) (formulating this test initially), *vacated*, 139 S. Ct. 2484 (2019); see also *League of Women Voters of Mich. v. Johnson*, 352 F. Supp. 3d 777, 804 (E.D. Mich. 2018) (adopting the test of the *Common Cause v. Rucho* court), *rev'd*, 902 F.3d 572 (6th Cir.).

²²⁵ See sources cited *supra* note 222.

²²⁶ Brief for Appellees League of Women Voters of North Carolina *supra* note 222, at 50–51.

²²⁷ Brief for Appellees, *supra* note 222, at 26–28. In particular, they have formulated the three prongs as:

1. Did the State consider citizens' protected First Amendment conduct with an intent to burden those citizens because of their political beliefs?
2. If so, did the redistricting map, in fact, dilute the votes of the targeted citizens or disrupt their political association in a discernable and concrete way?
3. If so, is there a constitutionally acceptable explanation for the map's ill effects, independent of the intent to discriminate on the basis of political belief?

Id. at 27–28.

²²⁸ Brief for Appellees League of Women Voters of North Carolina, *supra* note 222, at 52–55.

²²⁹ *Id.* at 55–60.

²³⁰ *Id.* at 60–63.

discriminatory effects but that is adopted without the requisite intent²³¹ would still have the sorts of pernicious effects that we might want to ban.²³² While the step test may increase the perception that these claims are “manageable” because the three prongs are “limited and precise,” this shift might come at the cost of considering the true nature of the underlying wrong.²³³

Neither version of the three-prong test convinced a majority of Justices that it offered workable methods to adjudicate partisan gerrymanders.²³⁴ Nevertheless, their success in the lower courts reveals something interesting about the intellectual hold that these multistep tests have over judges. Although it is impossible to prove that the multistep nature of these tests is causing judges to adopt them as potentially manageable standards for evaluating partisan gerrymandering claims, their form is suggestive. Indeed, Justice Elena Kagan’s dissent in the case highlights how familiar these sorts of three-step tests are in the law.²³⁵ Although advocates of a multistep test were ultimately unsuccessful in *Rucho*, the cases demonstrate the power of the rhetoric of stepification.

* * *

My goal in these examples is to provide instances of a process in action and to suggest that important commonalities can be observed in cases of stepification. While these are only four examples of stepification (some of

²³¹ Lower courts have required not only intent to discriminate but for that intent to actually predominate. *See* *Common Cause v. Rucho*, 318 F. Supp. 3d 777, 852 (M.D.N.C. 2018), *vacated*, 139 S. Ct. 2484 (2019).

²³² To be sure, this is potentially controversial and the intuition turns on what one’s view is of the reason gerrymanders are bad. If the issue is the intent of the legislature to “rig the game,” then intent is a “requirement” of an impermissible gerrymander. If, instead, the issue is certain forms of distortion per se, then perhaps a more flexible test would be beneficial.

²³³ Of course, for reasons of error costs, we might accept that the test will not cover all the situations we may actually want as a concession to the limits of the institutional competence of the courts.

²³⁴ *Rucho v. Common Cause*, 139 S. Ct. 2484, 2506–07 (2019). Though, it is worth noting that there is ongoing state litigation that will still make these issues salient, even if they are rejected at the Supreme Court. For example, plaintiffs made similar arguments under the North Carolina constitution for why these gerrymanders are illegal, with success. *See* *Common Cause v. Lewis*, No. 18 CVS 014001, 2019 WL 4569584, at *1–3 (N.C. Super Ct. Sept. 3, 2019). Similarly, the Pennsylvania supreme court ruled that the state’s legislative map was unconstitutional based on the state constitution, but interestingly chose not to adopt a step test and rather stated that it could consider several factors in determining whether a given gerrymander was unconstitutional. *See* *League of Women Voters of Pa. v. Commonwealth*, 178 A.3d 737, 815–18 (Pa. 2018) (stating that the court would consider “(1) the population of such districts must be equal, to the extent possible; (2) the district that is created must be comprised of compact and contiguous geographical territory; and (3) the district respects the boundaries of existing political subdivisions . . . [and] divides as few of those subdivisions as possible”).

²³⁵ *Rucho*, 139 S. Ct. at 2516 (Kagan, J., dissenting) (highlighting that many tests take this three-step form).

which include a move away from multistep tests), I believe that the evolution of these regimes suggests that there may be common normative implications of stepification. Other examples, including the relaxation of the three-prong standing doctrine²³⁶ in the case of states²³⁷ and the development of the two-step strict scrutiny analysis,²³⁸ further support this general trend.

In particular, stepification, while separate from the rules-versus-standards debate, has the advantage of its rule-like rhetoric. It distills issues that may have moving normative parts into simpler legal inquiries. For this reason, stepification may seem attractive to those who observe that an area of the law is chaotic. Stepification seems to impose order upon the chaos. This has made stepification an important tool for advocates against partisan gerrymandering to combat the perception that adjudication of these claims will lead to chaos. But stepification also has important drawbacks, making its normative valence ambiguous. The next Part turns to this question.

IV. THE UNEASY CASE FOR (AND AGAINST) STEPIFICATION

This Article does not argue that stepified tests are necessarily unhelpful. In some cases, they offer helpful frameworks for decisions and increase clarity. However, these advantages should not be mistaken for a strong case that stepification represents a universal solution to the problem of legal decision-making. This Part aims to briefly discuss some of the benefits and costs of stepification and to suggest that, while a strong case can be made for it in some circumstances, the general case for it is unclear.

A. *The Case for Steps*

Stepification has normative benefits that are important to highlight, even as this Article suggests that they are not as compelling as they might initially appear. In particular, four arguments in favor of these tests present themselves: (1) their value as rhetorical rules, (2) the ability to make easy cases of law easy in practice, (3) their ability to organize legal argumentation, and (4) their ability to give direction to lower courts. Importantly, some of these benefits are very similar to the purported advantages of bright-line rules. Stepification, then, might be seen as a judicial technology that allows these benefits to be imported to situations in which a bright-line rule cannot

²³⁶ See *Lujan v. Defenders of Wildlife*, 504 U.S. 555, 560–61 (1992).

²³⁷ See, e.g., *Massachusetts v. EPA*, 549 U.S. 497, 519–20 (2007) (noting that states are entitled to “special solicitude” for the purposes of standing inquiries).

²³⁸ Adam Winkler, *Fatal in Theory and Strict in Fact: An Empirical Analysis of Strict Scrutiny in the Federal Courts*, 59 VAND. L. REV. 793, 800 (2006) (“Justice Douglas’s phrase caught on and eventually became increasingly formalized into a two ‘prong’ test now referred to as ‘strict scrutiny’ or ‘compelling interest’ analysis.”).

be formulated and indeed where the individual steps might be standards. Still, this Section charts how stepified tests achieve these benefits and suggests some important ways in which multistep tests may subtly differ from bright-line rules.

First, despite any misgivings about these tests, they have the potential to create the perception of order for litigants. Leaving aside whether or not stepification actually creates order in practice, these tests may enhance the legitimacy of the law by clarifying why a particular litigant won or lost and creating the impression that the law is rule-like.²³⁹ In contrast to more open-ended balancing, the winners and losers in a stepified regime are more easily able to discern the reasons for the verdict in any particular case. In particular, they should be able to easily point to the step (or steps) on which they lost.

To the extent that this enhances public opinion of the law and those who enforce it, this might be a benefit for the legal system in general, even if it comes at the cost of certain individual cases being decided differently. Indeed, in many areas, as recently shown in the partisan gerrymandering space,²⁴⁰ courts may be reluctant to intervene where they perceive that they will not be able to articulate reasons for their decisions that will be convincing to observers. To the extent that stepification resolves these issues, it may represent a useful judicial technology.²⁴¹

To be sure, enhanced legitimacy may be harmful if it is based on a false perception. In particular, overconfidence in the reliability of legal judgments may make the populace less likely to search for ways to improve the legal system.²⁴² In particular, this dynamic may take on a darker tone if judicial rhetoric exists only to provide cover for elites to engage in politics by other means at the expense of the larger populace.

²³⁹ See Scalia, *supra* note 128, at 1178–79 (“When a case is accorded a different disposition from an earlier one, it is important, if the system of justice is to be respected, not only that the later case *be* different, but that it *be seen to be so*. . . . Much better, even at the expense of the mild substantive distortion that any generalization introduces, to have a clear, previously enunciated rule that one can point to in explanation of the decision.”).

²⁴⁰ See, e.g., Brief for Appellants at 39–40, *Rucho*, 139 S. Ct. 2484 (No. 18-422) (citing courts’ “persistent inability to discern any manageable standards for adjudicating partisan gerrymandering claims” and arguing that therefore the claims should be nonjusticiable).

²⁴¹ We can and should question whether this perception that these tests offer neutral ways to articulate reasons is more context dependent than it may initially appear. It could certainly be true that the issue here is somewhat circular and that individuals have been conditioned to think of these modes of decision-making as more official than they really are.

²⁴² Cf. RICHARD ALBERT, CONSTITUTIONAL AMENDMENTS 162–64 (2019) (noting that in the context of constitutions, there is evidence to suggest “veneration” of the “constitutional status quo” makes individuals less likely to support amendments, though constitutional cultures do exist in which frequent constitutional change occurs and so a culture of reform becomes embedded in what individuals venerate about the constitution).

But the question of whether legitimacy is good or bad ultimately turns on whether the legal system itself is worthy of such veneration. Professor Tom Tyler has demonstrated that individuals often feel a strong obligation to obey the law even if they feel it is unjust.²⁴³ Further, the perception that a process has been procedurally fair is independently a source of legitimacy for the system outside of whether favorable outcomes for the individual occurred.²⁴⁴ To the extent stepification increases judgments of procedural fairness—a question perhaps deserving of separate study—then we should ask whether the underlying system is deserving of those benefits.

Second, these tests arguably help make easy cases easier, even if they struggle to appropriately capture the interesting dimensions of the more difficult questions of law.²⁴⁵ While the proportion of easy and hard cases is context dependent, the ability to invoke a particular portion of the analysis to explain why one party lost can potentially save time by creating conceptual clarity. While these tests arguably flatten and simplify certain areas of the law, this simplicity can sometimes be a boon for judges working to resolve cases that are more straightforward.

Third, and perhaps more convincing, these tests provide useful ways to organize legal argumentation. A test that is divided into multiple clear steps gives parties and prospective litigants a natural way to structure their thinking on a given issue and to organize their briefing. While easing the work of lawyers is not necessarily a reason to alter the structure of legal rules, the benefit to regulated parties may still be substantial if the clarity offered by a step test offers them a way to structure their behavior so as to avoid violating the law.

This third benefit is tied to the first one. To the extent that briefing aligns with the actual decision process of the court, rhetorical efficiency can be achieved. For example, say the parties know that three questions matter: the court can reply that, in fact, the plaintiffs won because of the result of question two and organize the analysis to this effect. This means that courts can also decline to address questions that they do not want to answer by appealing to other parts of the step test, an option not available under a

²⁴³ TOM R. TYLER, WHY PEOPLE OBEY THE LAW 46 (1990).

²⁴⁴ *Id.* at 77–79.

²⁴⁵ Professor Kaplow makes a similar point, though he notes that in cases where the analysis terminates it does so because the answer is wrong. Kaplow, *On the Design of Legal Rules*, *supra* note 22, at 1000–03. This is not the case for general multistep tests because steps one and two are not necessarily balanced against each other. Thus, the law can fail to be clearly established and the plaintiff fails without the test resulting in an error.

balancing approach.²⁴⁶ This can also help limit the insertion of dicta into opinions.²⁴⁷

Fourth and finally, step tests provide appellate courts with a method to structure and monitor the actions of lower courts. By forcing courts to articulate reasons in this form, the reviewing court also gains some ability to oversee the actions of the lower courts.²⁴⁸ A higher court can perhaps more easily detect errors or areas of disagreement because it will be able to identify the part of the test on which it disagrees with the lower court.²⁴⁹ More directly, stepification can make the decisions of lower courts more legible by organizing their reasoning in a way that is easier for appellate courts to monitor and evaluate and to which parties can point when identifying errors.²⁵⁰ To the extent that the appeals process exists to discipline lower court discretion, stepification may further enable that process.

It is possible to sketch a model of this dynamic mathematically by comparing a three-factor balancing test to a three-prong step test. In each balancing decision, the appellate court will compute the three factors and convert these three results into a final outcome in the case. The lower court will then have guidance in two ways. First, it can compare the case before it to the appellate decision on each of the three factors. Second, if the case before the lower court is somehow “stronger” on each factor then the result

²⁴⁶ See, e.g., *Burwell v. Hobby Lobby Stores, Inc.*, 573 U.S. 682, 727–28 (2014) (declining to adjudicate whether the contraceptive mandate served a compelling governmental interest by concluding that the action violated the second step of the test requiring that it was not the least restrictive means to serve that interest).

²⁴⁷ I do not mean to say this will always be the case. For a court to explain why a party satisfies all the steps it must indeed discuss all steps the parties contest. In contrast, it might be that a party makes such a strong showing on one factor of a multifactor balancing test that the other factors are rendered irrelevant. Indeed, Professor Barton Beebe has argued that in the context of multifactor test for trademark infringement, judges will make a small number of factors dispositive and then stampede quickly over the rest. Barton Beebe, *An Empirical Study of the Multifactor Tests for Trademark Infringement*, 94 CALIF. L. REV. 1581, 1586–87 (2006).

²⁴⁸ This might also take the form of limiting further doctrinal development on top of the stepified test. See Michael Coenen, *Rules Against Rulification*, 124 YALE L.J. 644, 658–60 (2014) (discussing the notion of rules against rulification, in which an appellate court forbids lower courts from attempting to convert a standard into a rule). While Professor Coenen’s analysis focuses on the building of further rules on top of an existing framework, we can see here a desire for the courts to preserve the structure of an existing regime rather than to allow deviations.

²⁴⁹ Cf. Kaplow, *On the Design of Legal Rules*, *supra* note 22, at 1050 (arguing that in cases where both balancing and structured processes require the reviewer to assess harms and benefits against a threshold, it does not seem as though it would constrain discretion). Where the test focuses on questions of law rather than fact, the result may be different.

²⁵⁰ Cf. JAMES C. SCOTT, *SEEING LIKE A STATE* 2–6, 183–84 (1998) (examining how the notion of legibility and how the state can or cannot order affairs to enable those affairs to be tracked by the state both facilitate administration and can lead to failure).

will be clear.²⁵¹ If, however, a comparison of the factors produced a more complicated result, the lower court will have to choose how to resolve the case.

In contrast, for each multistep case (where the court completes the steps), the lower court is given guidance on each of the three steps, all of which are potentially dispositive about the case. The lower court similarly knows that if its case has the same combination of answers to the steps, then the result will be clear.²⁵²

While an individual multistep decision and individual balancing decision have a similar result on instructions to the lower court, as the appellate decisions add up, the multistep decisions have greater cumulative effect. Each multistep decision provides a new boundary for each individual step, each of which can be dispositive.²⁵³ In contrast, each balancing decision provides a new input on each factor but can still leave it unclear how novel combinations of factors should be evaluated. The result is more openness to individualized judgment when such new combinations present themselves.

Channeling lower court discretion in this way potentially leads to greater uniformity in the law in the sense that lower courts apply the law in the way the appellate court would.²⁵⁴ This is not to imply that lower courts are attempting to subvert the will of other tribunals. But a multistep test allows precedent to build around individual questions. Thus, the Supreme Court can issue opinions on what it means for the law to be “clearly established” for the purposes of qualified immunity, for example. In contrast, under a balancing regime each precedent tells us less about the interplay of factors when reaching the final result, offering less guidance to lower courts.

Note that the dynamics of appeal matter here too. Appellate tribunals also provide guidance to lawyers looking to advise their clients. To the extent

²⁵¹ We can think of the appellate decision as coding a case in three-dimensional space where each axis represents a factor. This coding not only tells us about this point but also tells us about the “cone” of points that are more extreme than it is on all three factors. Still, if a new case presents a stronger showing on two factors but a weaker showing on one factor, then the appellate case is not strictly dispositive of the result, though it may be close enough that the court can reason by analogy.

²⁵² In this sense, the individual multistep decision cuts out the same cone because it sets up three inequalities in the three-dimensional space. This is because on each individual question, the lower court is given a single answer—assume it is “yes”—and can then conclude that anything more extreme has the same answer. It cannot, however, necessarily determine in the other direction when yes would become no.

²⁵³ Assuming, that is, that the court actually proceeds through all the steps.

²⁵⁴ Cf. Coenen, *supra* note 248, at 685 (noting that restrictions on imposing structure on mandatory standards limits doctrinal experimentation). Professor Peter Strauss also has argued that *Chevron* may function as a management tool to increase the Supreme Court’s control over lower courts. Peter L. Strauss, *One Hundred Fifty Cases per Year: Some Implications of the Supreme Court’s Limited Resources for Judicial Review of Agency Action*, 87 COLUM. L. REV. 1093, 1095 (1987).

that multistep tests make errors more legible to the appellate court, they also make them more legible to lawyers who can provide their clients with more definitive advice about their chance of success in a costly appeal. “The court got it wrong on step one” is, I believe, more definitive than “the appellate court will likely balance the factors in our favor.” That said, I don’t want to overstate this point. Some analysis of multifactor balancing tests in practice do find that judges zero in on a few factors.²⁵⁵ But even if this is so, the question still remains whether the choice of factors is consistent or whether the multistep test still produces more coherence.²⁵⁶

While these benefits seem largely procedural, they are arguably more substantive than they initially appear. Consistency does matter in the law and, as such, so do improvements in the ability of the parties and courts to focus themselves on the right questions. At the same time, there are reasons to believe that uninterrogated acceptance of these tests has normative costs, or at least important implications.

B. *The Trouble with Stepification*

Despite the advantages of stepification detailed above, it is by no means a blind endorsement of the step regime. In particular, we should be skeptical of an unthinking acceptance of stepified tests as a superior option to balancing tests. My critique here applies to the general category of step tests rather than any individual one,²⁵⁷ and therefore it necessarily sweeps fairly broadly. We can identify three issues with the general trend of stepification. The first argument against these tests is that they flatten certain normative elements of the law into a more rigid, procedural structure, potentially obscuring important elements of the issue itself. This leads to several related problems, including the bleeding of the supposedly isolated prongs of the test into each other and the tendency to create complexity over time. Second, they may impart artificial simplicity to decisions. Third, they may overly

²⁵⁵ See Beebe, *supra* note 247, at 1586 (“[J]udges employ *fast and frugal* heuristics to short-circuit the multifactor test. Perhaps as an expression of their cognitive limitations, but more likely as an expression of their cognitive ingenuity, judges rely upon a few factors or combinations of factors to make their decisions. The rest of the factors are at best redundant and at worst irrelevant.” (footnote omitted)).

²⁵⁶ Professor Beebe’s analysis is interesting on this point. He does find that the circuits themselves seem to have standards about which factor matters most but that the answer varies by circuit. *Id.* at 1598–99. In this case, there is partial uniformity within circuits but not national uniformity. *Id.*

²⁵⁷ Of course, a wealth of critiques of individual stepified regimes exists. See, e.g., Stephen Breyer, *Judicial Review of Questions of Law and Policy*, 38 ADMIN. L. REV. 363, 373 (1986) (critiquing an overbroad reading of *Chevron*); Paul W. Hughes, *Not a Failed Experiment: Wilson-Saucier Sequencing and the Articulation of Constitutional Rights*, 80 U. COLO. L. REV. 401, 428 & n.121 (2009) (arguing that the optional sequencing of the two steps of the qualified immunity analysis inquiry decreases the rate of constitutional adjudication).

constrain lower courts in the development of the law. Thus, the drawbacks of stepification largely mirror its advantages.

1. Flattened Normative Dimensions

The most important objection to stepification is that it presumes an orthogonal structure to the decision at hand and sometimes also a lexicographical ordering of the issues. Both of these assumptions, while defensible in some cases, are normatively problematic in others. First, stepification assumes that some issues can be cleanly separated from others. That is, it is possible to reduce the overall issue to several logically separate prongs that can be evaluated without reference to the other inquiries. Second, stepification, by ordering the issues in question, can implicitly assume the priority of one over the other, both normatively and practically.

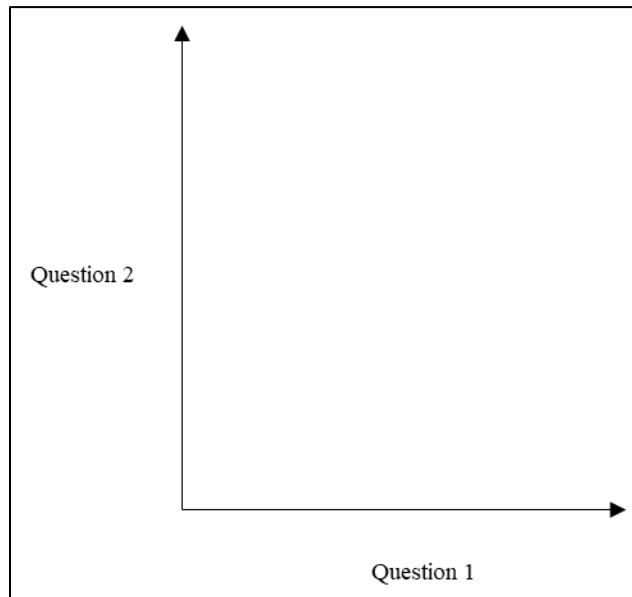
To deal with the former assumption first, stepification requires that the issue in question—the one the test is designed to resolve—can be severed in a particular way. That is, stepification of the normative issue the law attempts to regulate via a multistep test implicitly assumes that the factors captured by the prongs can be normatively evaluated separately. Alternatively stated, the multistep form presumes that the *degree* to which a party prevails is irrelevant to whether she ought to prevail on the subsequent steps. But whether such walls exist is a normative question, and where the walls do not obviously exist a normative judgment is imposed.²⁵⁸ In the case of traditional multielement offenses, this requirement may be satisfied. However, in issues with more complicated normative dimensions, this feature should not be taken for granted.

To make this observation more concrete, we can consider a graphical representation of stepification. Multistep tests assume the issues can be separated from each other for the purpose of decision-making. It may be true that the way in which a previous step is resolved will influence the starting point for later steps, but the results of previous steps of the inquiry will be taken as a given for the purpose of subsequent analysis. As a result, multistep

²⁵⁸ Another way to state this is to say that where an issue is “modular” it can be more easily rendered into a multistep test. See Henry E. Smith, *Property as the Law of Things*, 125 HARV. L. REV. 1691, 1701 (2012) (describing property law as a modular system in which “chunks or components of the system to be partially walled off and the interconnections between these chunks and the rest of the system to be deliberately limited (sometimes even at the expense of interdependencies that might have some value”). Professor Smith argues that modularity is then a way of managing complexity as interactions between components are limited. *Id.* Professor Smith’s work has an important echo in some of the mathematical language used here about orthogonality. At a high level of generality, some mathematical structures can be decomposed into smaller components that can be then examined individually. Courts imposing multistep tests on an area of doctrine are doing something similar in my telling: while a single rule or standard might not fit the problem, a sequence of them might. This modularity can then manage complexity in the way Professor Smith describes.

tests essentially posit that the inquiries are orthogonal with one issue unrelated to each other, creating the framework in Figure 3:

FIGURE 3: ORTHOGONAL QUESTIONS IN A TWO-STEP TEST



We can then depict a hypothetical range of outcomes that pass the test and others that do not. For example, a situation that might be more amenable to balancing is depicted in Figure 4:

FIGURE 4: MAPPING A TWO-STEP TEST TO OUTCOMES

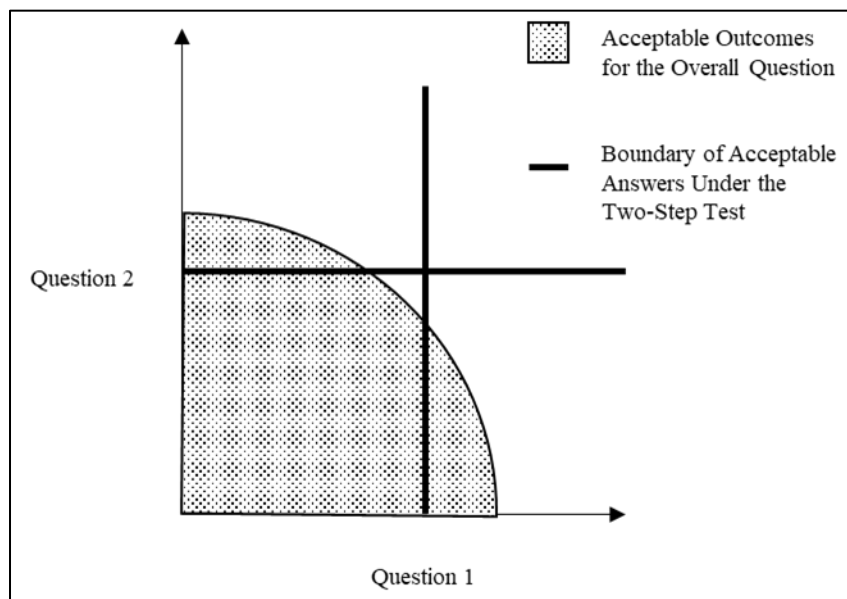


Figure 4 presents a highly stylized picture of the issue with stepification.²⁵⁹ Because the decision-maker is forced to answer each question separately, they must draw individual lines for each question. The result is that the resulting space of outcomes that pass the test may not match the underlying desired results, excluding some outcomes that we might wish to admit and allowing certain actions that might be undesirable.²⁶⁰ For example, consider the form of the qualified immunity test.²⁶¹ If we assume the test is intended to protect officers who act in good faith, the test can be

²⁵⁹ Note that the lines in question must be orthogonal to the axis they correspond to for each question to be truly independent of the answer of the former. One could imagine them bending together or being slanted, but this would be to allow the two questions to bleed into each other. As discussed later in this Section and in *infra* Part V, I believe that in practice courts are forced to bend the lines to accommodate cases that stretch the limits of the test, but to do so is to deform the values of stepification. Of course, the end result might be a better fit for the landscape and therefore may produce superior results. It only does so, however, by reproducing the balancing analysis that stepification is designed to replace.

²⁶⁰ One way to think about this issue is that the process of jurisprudence has elements that are topological in the mathematical sense. That is, the formulation of legal rules (whether done by courts or the legislature) is done to both reshape and respond to the special qualities of the underlying issue. Sometimes, the issue will have a structure that admits a particular jurisprudential approach (whether a bright-line rule or a multistep test) and in other cases it will not.

²⁶¹ See *supra* note 56 and accompanying text.

both underinclusive and overinclusive.²⁶² It might protect officers who act in bad faith where the law is not clearly established while also offering no protection to officers acting in good faith where there is a precedent in a complicated area of law that the officer is unaware of—perhaps because the decision is recent.

The preceding analysis highlights a further point: the degree to which a multistep test fails depends on the underlying value it is intended to capture. Importantly, if the text is identified with the value it aims to capture then there will be no imperfection. For example, if we decide that officers should only be liable for clearly established constitutional violations, then qualified immunity will be a perfect fit for the problem. Like many doctrinal formulations though, multistep tests will frequently fit the underlying subject matter imperfectly. In particular, where at the margin the factors captured by the steps should be balanced, a multistep test will fail to always reach the right results.²⁶³ Still, multistep tests may offer a way to simplify decisions in the majority of cases, but issues of fit may plague the boundaries.

A concrete example helps illustrate the point. Consider the gerrymandering cases discussed above. The proposed tests purported to limit the need for court involvement by using individual prongs to carve out categories of gerrymanders that would be exempt from their coverage.²⁶⁴ While this approach is all well and good in the abstract, it seems to flatten some issues that we may find normatively significant. For example, what about a map that was made with purely invidious intent but is one among many that could be explained by the state's geography? Perhaps as a pragmatic matter, this is an inquiry that is better suited to political resolution, but as a normative matter, is it any less offensive to the democratic process? Similarly, would a map with dramatic partisan effects, no explanation from the state's political geography, but no evidence of intent be inoffensive to constitutional values (if not doctrine)?²⁶⁵ No matter how well suited to a

²⁶² This may be a controversial account of what the modern test is intended to protect. The modern test might serve other goals. I choose this value here both to illustrate the point as well as because it is rooted in the history of the test. That is, the modern qualified immunity test replaced the prior standard that explicitly based immunity in good faith because the Court found that explicit consideration of subjective motivations allows insubstantial claims to proceed to trial. *Harlow v. Fitzgerald*, 457 U.S. 800, 815–17 (1982).

²⁶³ Kaplow, *On the Design of Legal Rules*, *supra* note 22, at 1050–55.

²⁶⁴ See *supra* Section III.D.

²⁶⁵ Note that I do not mean to suggest that current doctrine would take this approach, and it is well-known that there are no disparate impact claims under the Fourteenth Amendment. *Washington v. Davis*, 426 U.S. 229, 246–48 (1976). My point is that an ideal legal prohibition (perhaps crafted by the legislature) might rule out various maps that cannot meet all three prongs of the suggested test where they are sufficiently offensive to an individual one. And they might do so because the individual steps are not

given problem the proposed tests were, by separating the elements of the inquiry, they overlook critical interactions. In particular, because an issue must pass each step of the test individually, cases can occur that barely satisfy the test but which offend the underlying value the test aims to protect.

Going further, the *Lemon* inquiry has arguably broken down because its factors are not approximately independent of each other.²⁶⁶ If there is an entanglement between the different elements of a given issue, the steps of the test erase that relationship. Because a court is bound to apply every step individually, it cannot look at the nexus of the issues in question. Thus, when cases that are a poor fit for the test are presented, they force adjustments or abandonment of the test. While *Lemon* may resolve many of the core cases to which it is intended to be applied, the issues at its boundaries may create long-run issues for its viability. For example, the issues in *Van Orden* and the more recent *American Humanist Association* presented the Court with cases in which *Lemon* seemed to produce the wrong result in the minds of a majority of the court.²⁶⁷

One solution here is to let the factors bleed into each other, balancing one against the other to compensate for these issues. But this solution essentially sacrifices the desirable property of the step test that keeps these issues separate. Once one accepts that the issues are not entirely separate, the test begins to collapse towards balancing once more, undermining the viability of the entire enterprise. While the new modified test might retain some of the rhetorical benefits of stepification, it would forfeit the intellectual advantages stepification is supposed to create.²⁶⁸

Of course, stepified regimes are not always destined to collapse because of some internal tension between the factors. Again, with certain issues, the separation between issues is clear. And, in general, situations that mirror the multielement offenses found in criminal law that require a showing of both the requisite *mens rea* and *actus reus* seem naturally fit for stepification. Though even here, the criminal law accomplishes this natural separation by grading crimes rather than by applying a simple two-step test.²⁶⁹

necessarily akin to on–off switches but capture different values about how districts should be drawn, not all of which are necessary to conclude that a given map offends democratic values.

²⁶⁶ See *supra* Section III.B.

²⁶⁷ See *supra* notes 192–199 and accompanying text.

²⁶⁸ See *supra* Section IV.A.

²⁶⁹ For example, a pure stepification approach would look first to some level of *mens rea* and then look for the *actus reus*. But the criminal law realizes that this sort of yes–no approach would risk treating very different normative situations as identical crimes. Therefore, the *mens rea* question will often determine the *grade* of the crime rather than determining whether there is a crime at all. This use of the question for tiering is distinct from stepification because it departs from the usual yes–no structure of the

Instead, the central problems arise in areas in which the creation of steps serves to flatten an issue into artificial inquiries that fail to capture the complete picture. For example, consider Justice Kavanaugh's proposed refinement of *Chevron*.²⁷⁰ Justice Kavanaugh's alternative still features steps, but he alters the nature of the inquiry to focus on the issues he believes are more naturally handled by judges.²⁷¹ Leaving aside whether this proposed change is an improvement over the current test, Justice Kavanaugh's proposal is revealing because it shows that the issue of deference to the executive branch is not so neatly diced into any particular categories. Rather, the nature of the issue at hand as well as its complexity admit alternative categorizations of the issues. In the language of management discussed above, an alternative decision tree may be posited to resolve the issue.²⁷² Thus, any specification of the issues in a stepified form likely requires a choice among alternatives that will have normative implications for the functioning of the test.

This point has two main implications. First, stepification may lead to errors. Within the rigidity of a given regime, there might be cases that narrowly answer each element of the inquiry without actually satisfying the underlying normative goal of the legal regime in question. In some sense, they would be right on each part of the test, but wrong when looked at as a whole. Indeed, this is one way to characterize major questions doctrine cases. In these cases, the agency interpretation literally passes the test but might seem to be doing so in some way that violates some other background principle of the limits of agency power to decide, well, major issues. While certainly debate can be had over the result of individual cases and their doctrinal implications, one way to read them is as an admission that the sum of the parts can be wrong in a way that the individual parts were not.²⁷³ If stepification leads to issues falling through the cracks in this way, we might look less favorably upon it.

Indeed, Professor Kaplow argues that the structured decision procedures he studies have precisely this effect. In cases where they actually

typical multistep test. That said, as many students of criminal law have realized over the years, this process can be helpfully expressed in a decision-tree framework. Thus, an approach mirroring the use of decision frameworks in criminal law is one way I suggest courts could move beyond stepification. *See infra* Section V.B.

²⁷⁰ *See supra* notes 179–183 and accompanying text.

²⁷¹ Kavanaugh, *supra* note 11, at 2154 (“But in cases where an agency is instead interpreting a specific statutory term or phrase, courts should determine whether the agency’s interpretation is the best reading of the statutory text. Judges are trained to do that, and it can be done in a neutral and impartial manner in most cases.”).

²⁷² *See supra* Section II.B.2.

²⁷³ *See supra* notes 176–178 and accompanying text.

reduce effort, they do so by reaching the wrong answer.²⁷⁴ In all other cases, the court must proceed through the different steps regardless.²⁷⁵ There is no real efficiency gain because simplification only leads to errors.

Second, because complex issues do not admit stepification in some natural way, alternative step tests can be formulated that may lead to different results. In such cases, the choice of test itself will have important normative implications. The decision will not end at whether to choose a multistep test but will extend to which test to use. And the choice of test itself might have implications for how the inquiry is structured and the subsequent system that is built up around that choice.

To illustrate this point, consider the steps proffered in the gerrymandering cases discussed above. The choice of the first step—whether there is intent²⁷⁶—is a normative one. We can imagine a test that does not look at intent but instead only assesses impact. We could also imagine a test that starts with impact before moving on to other considerations. But by designing a step test that, first, makes intent a consideration and, second, makes it the *first* consideration (this has important lexicographic implications discussed below), courts implicitly impute certain value judgments into the step test. In this case, a value judgment implicit in this test is that the *primary* ill in gerrymandering is the intent, not the practical impact on the ability of citizens to participate in competitive elections. Therefore, with issues that have a deep moral valence, not only does deciding whether to adopt a step test import normative judgments, but also which steps to use carries important normative implications.²⁷⁷

One can see a related issue in the effort of the Supreme Court to work through the implications of its decision in *Miranda v. Arizona*.²⁷⁸ *Miranda* is now known for its famous warnings, not as a multistep test. The opinion itself, however, implicitly set up such a test that was later taken up by lower courts, requiring the warnings to be issued in the case of “custodial interrogation.”²⁷⁹ But then one must ask: what are (1) custody and

²⁷⁴ Kaplow, *On the Design of Legal Rules*, *supra* note 22, at 1002.

²⁷⁵ *Id.* at 1000–03.

²⁷⁶ *See supra* note 227 and accompanying text.

²⁷⁷ To be sure, doctrine may be an imperfect expression of underlying moral concerns. I therefore do not want to imply that the choice of doctrinal form necessarily implies that advocates (or courts) automatically believe that intent is the primary issue with gerrymandering whenever they suggest this test. But even so, there is still a message that is expressed by the choice of the steps, even if the test’s creators or advocates privately think otherwise.

²⁷⁸ 384 U.S. 436 (1966).

²⁷⁹ *Id.* at 444 (“By custodial interrogation, we mean questioning initiated by law enforcement officers after a person has been taken into custody or otherwise deprived of his freedom of action in any significant way.” (footnote omitted)).

(2) interrogation. This sets up a two-step test, though the steps are unordered. A substantial amount of work, however, must go into interpreting those factors. In some cases, there is no custody and so the warnings do not apply.²⁸⁰ In other cases, the court needed to interpret the meaning of interrogation, sometimes introducing new branches of the analysis²⁸¹ and other times introducing exceptions to situations that passed the literal test.²⁸² Thus, where a multistep test is formulated to organize the law around a complicated issue, it may require appellate courts to build out ever more multistep tests within individuals steps of the original formulation, necessitating once more the choices I argue are occasionally fraught.

The second, related difficulty with stepification is its lexicographic ordering of issues. Mathematically, a lexicographic ordering of issues is one in which a given dimension of an issue is deemed more important than another for the purposes of ordering. Thus, that dimension is evaluated first to resolve any decisions, and only if it fails to resolve the issue are further dimensions considered.²⁸³

Step tests often seem to imply that one issue has priority over another—though sequencing is sometimes optional.²⁸⁴ In some cases the mandatory sequencing of the steps is a product of logical necessity. For example, there is often an implicit need in the law for an analytical step zero.²⁸⁵ However, that need should not be mistaken for the conclusion that steps are more generally neutral in their order. Rather, the separation of issues into a given order creates (and may imply that the issue implicitly has) a structure that has important consequences for the law in general. Thus, even in cases where the steps could theoretically be addressed in either order, the choice to order the issues in one way or another shapes how subsequent decision-makers will approach the problem. For example, in the gerrymandering example discussed above, the choice to address intent first may lead to fewer decisions that interpret the impact prong if courts skip over this analysis after

²⁸⁰ See *Berkemer v. McCarty*, 468 U.S. 420, 429 (1984) (holding that traffic stops did not constitute custody for *Miranda* purposes).

²⁸¹ See *Rhode Island v. Innis*, 446 U.S. 291, 300–01 (1980) (“We conclude that the *Miranda* safeguards come into play whenever a person in custody is subjected to either express questioning or its functional equivalent.”). This sets up an additional step in the analysis. After the custody question there are now two yes–no ways in which someone can be interrogated.

²⁸² See *Pennsylvania v. Muniz*, 496 U.S. 582, 601 (1990) (holding that a subject asked booking questions was interrogated for the purpose of *Miranda* but that these questions fell into a “routine booking question” exception).

²⁸³ A lexicographic ordering can order further dimensions as well. The second can be deemed superior to the third, the third to the fourth, and so on.

²⁸⁴ For example, this is the case in the qualified immunity context. *Pearson v. Callahan*, 555 U.S. 223, 234–36 (2009) (holding that lower courts may address the second step of the test first).

²⁸⁵ See *supra* Section I.A.

the plaintiff fails on the first step. Moreover, this property has significant implications for step tests, including how they evolve to compensate for deficiencies in their application.

Most important, steps can force decisions that might not otherwise be made or even evade decisions altogether. For example, Professor Re observes that a two-step *Chevron* test forces courts to opine on both the ambiguity of a term as well as the interpretation the agency offers.²⁸⁶ Similarly, the breakdown of the mandatory sequencing of the qualified immunity two-step test has had implications for how courts evaluate such claims.²⁸⁷

While step tests need not resolve all issues at once, or even in a given order, the fact that they often *do* order decisions in this way is still significant. It matters whether the court reaches certain issues as does the structure of questions.²⁸⁸ Which issues must be addressed first informs which issues will bind future courts and which will inform legal scholarship.²⁸⁹

Of course, certain issues necessarily have a logical structure of priority. To point to ordering as an issue that arises in stepification is not to say that ordering issues are per se negative. The problem occurs when ordering can obscure important dimensions of an issue in the process of attempting to formalize the law in question. Step tests do not have to be ordered (and in fact in some cases they break down by eliminating order),²⁹⁰ but where they are, the order in question has important implications.

Perhaps the paradigmatic example of these issues is *Chevron*. One way to state the goal of the doctrine in the abstract would be as follows: statutes often leave certain issues underdetermined and, within reason, the Executive Branch should be allowed to fill those gaps in the statutory scheme. What then is the merit of the two-step test? In the language of my thesis, it creates a useful framework for implementing an underlying view about the propriety of agency statutory interpretation, but its model is imperfect. In particular, it has the fitting problem that is depicted in Figure 4, in the sense that a literal application of the test seems to allow certain actions that seem like they

²⁸⁶ Re, *supra* note 59, at 608 (“On reflection, there are important advantages and disadvantages to traditional *Chevron*’s command that courts should ask about both mandatoriness and reasonableness in every case. For example, requiring courts to answer both questions facilitates the rapid development of the law, but asking only about reasonableness seems consistent with principles of judicial restraint.”).

²⁸⁷ Hughes, *supra* note 257, at 428 & n.121.

²⁸⁸ Re, *supra* note 59, at 608.

²⁸⁹ For example, if Step One of *Chevron* resolves most of the issues, commentary is sure to focus on step one and what will meet that burden.

²⁹⁰ See, e.g., *Pearson v. Callahan*, 555 U.S. 223, 234–36 (2008) (relaxing the rigid framework in favor of a more flexible inquiry).

should be restricted.²⁹¹ Thus, the test must be bent and modified to cover its deficiencies. The factors must be allowed to bleed into each other—akin to allowing the lines in Figure 4 to bend—or additional steps must be added over time to compensate for its issues—akin to adding additional lines. In essence, this description can be seen to generalize the issues that have plagued *Chevron* to the general setting of multistep tests. This description also shows how multistep tests can encounter difficulties over time.

To some, these remarks may seem somewhat overwrought, focused on abstract issues of form rather than the substance of the issue in question. Yet form matters.²⁹² While much can be said about the individual doctrinal tests surveyed, the point of this Article is that there is a thread connecting the issues in question. Steps both separate issues and order them. Sometimes this shift is useful for courts, but in other cases the separation and ordering have unintended consequences because of their imperfect fit with underlying issues. Most important, the specific way in which multistep tests struggle has important implications for how they eventually break down. The particular issues they encounter lead courts to modify them in ways that affect their ultimate form over time, particularly when the test addresses a controversial issue.²⁹³

2. *Easing Decisions*

As a second matter, even if stepification does not lead to a proliferation of errors, we might wonder whether there are costs to the *appearance* of ease

²⁹¹ For example, Professors Shoba Wadhia and Christopher Walker argue that the rationale for *Chevron* is weakest in immigration enforcement. Shoba Sivaprasad Wadhia & Christopher J. Walker, *The Case Against Chevron Deference in Immigration Adjudication*, 70 DUKE L.J. 1197, 1201 (2021). In such cases, they argue that even if the agency reaches an interpretation of the law through adjudication that can satisfy *Chevron*, the traditional rationale for deferring to such agency interpretations is particularly weak. *Id.* at 1201–02. They call instead for *Skidmore* deference in such cases. *Id.* at 1202–03. Though their argument against *Chevron* in such contexts is procedural, not substantive, *id.* at 1202, for this proposal to have an impact it will have to impact cases in which the result from applying *Chevron* differs from that which would be reached by balancing in the way Figure 4 illustrates. In other words, the procedural considerations the Professors identify give us reasons to not defer in cases where *Chevron* is overinclusive or underinclusive. For a different example, opponents of qualified immunity often point to cases where the doctrine shields officers from liability even where the facts of the constitutional violation seem egregious. *See, e.g.*, David Deerson, *The Case Against Qualified Immunity*, NAT'L REV. (July 13, 2020, 3:50 PM), <https://www.nationalreview.com/bench-memos/the-case-against-qualified-immunity/> [<https://perma.cc/5UDM-9CQQ>]. In doing so, these critics are not arguing that the court in question applies the doctrine wrong, but that the application of the current formulation is flawed because its *proper* application still leads to certain bad outcomes at the extremes.

²⁹² For example, Justice Scalia's famous case for preferring rules over standards is an example of how the Justice considered form to be important even abstracted away from the substance of individual cases. Scalia, *supra* note 128, at 1178–79. For a recent project arguing that legal doctrine and the way in which it is formulated is itself worthy of study, see SCHLAG & GRIFFIN, *supra* note 21, at 4.

²⁹³ *See infra* Part V.

in certain cases. Consider again the issue in *Brown & Williamson*: if the Court had concluded that the agency interpretation satisfied *Chevron*, would this necessarily be a satisfying reason to accept such a dramatic change to the law? Even accepting the wisdom of *Chevron*, there is perhaps something unsatisfying about explaining the solution to a “major question” as a matter of a clear two-step process. To the extent that stepification makes easy cases easy, it also can make hard cases (seemingly) easy as well,²⁹⁴ perhaps at the cost of the ability to persuade observers.

To suggest that such a rhetorical simplification of complex issues has normative implications is not novel. In his *Harvard Law Review* foreword, Professor Dan Kahan has observed that judicial opinions are often problematically simplistic.²⁹⁵ Professor Kahan observes that, contrary to certain intuitions, such certainty is deeply problematic.²⁹⁶ Indeed, certainty can deepen conflict rather than alleviating it.²⁹⁷

While this is not an Article about writing opinions, Professor Kahan’s conclusions apply here as well. Stepification simplifies areas of the law and, in doing so, simplifies the opinions that are issued by courts deciding those issues. To the extent that Professor Kahan is correct about the limitations of this sort of certitude about the law, the articulation of simple solutions to complex issues may have unintended drawbacks for decision-makers looking to persuade observers. In particular, opinions that are too neat and tidy might make certain litigants feel as though their concerns have been dismissed. Thus, the simplification offered by multistep tests might be both a virtue and a vice.²⁹⁸

Thus, even though something is potentially gained through stepification, there might be important normative losses to this sort of rigidity. These losses may lead to errors in substantive cases, but they also may appear when we consider the process value of reason-giving. The cost of neat conceptual boxes is that they might sometimes obscure more than they reveal. Stepification is no different.

²⁹⁴ See *supra* Section IV.A.

²⁹⁵ Dan M. Kahan, *The Supreme Court 2010 Term—Foreword: Neutral Principles, Motivated Cognition, and Some Problems for Constitutional Law*, 125 HARV. L. REV. 1, 59 (2011) (“Judicial opinions are notoriously—even comically—unequivocal. It is rare for opinions to acknowledge that an issue is difficult, much less that there are strong arguments on both sides.”).

²⁹⁶ *Id.* at 60.

²⁹⁷ *Id.* (“But in fact, the opposite is more likely true. Studies of motivated cognition and related dynamics show that pronouncements of *certitude* deepen group-based conflict.”).

²⁹⁸ See *supra* notes 239–241 and accompanying text (discussing the potential for an opposite effect). In particular, the effect of simplified, more rule-like opinions that might result from stepification is uncertain and perhaps worthy of empirical study.

3. *Excessive Control of Lower Courts*

Finally, it is worth touching upon a downside associated with one of the aforementioned advantages of stepification.²⁹⁹ While stepification is a potent way to control lower court discretion, it may impede the development of diversity in the lower courts, controlling them too much. That is, the inverse of the observation about the benefits of control can also be made.

In some cases, we might *prefer* nonuniformity at the lower court levels. More amorphous tests may allow local judgments or norms to creep into judgment in a positive way. For example, preserving probable cause as an inquiry that looks to the totality of the circumstances may enable local courts to make localized judgments free from rigid frameworks articulated by appellate courts.³⁰⁰ Without romanticizing localized judgments, there are many legal inquiries that benefit from freedom to tailor results to the community.³⁰¹ A more structured version of law is an advantage in certain situations. Yet this form of control has drawbacks. Resolution of cases at one step or another concentrates thoughts on those efforts rather than bringing the entire issue into focus.

This point is susceptible to the counter that amorphous balancing tests are no different for the purposes of developing the law. To the extent that cases involving balancing are read broadly or used to reduce the inquiry in practice to a single factor then the result may be no different.³⁰² But as long as judges are willing to read precedent as a guide for the application of principles and not just of a resolution to individual cases, there may still be an advantage obtained by balancing approaches. Courts that are freed of analytical shackles may develop parts of the law that will otherwise not be reached by more limited analyses, thereby spurring further development.³⁰³

²⁹⁹ See *supra* Section IV.A.

³⁰⁰ Cf. Coenen, *supra* note 248, at 689 (arguing that limits on adding structure to standards empower trial courts at the expense of appellate tribunals).

³⁰¹ For example, the totality-of-the-circumstances standard for informant tips may properly allow judges to consider local conditions more effectively when considering whether tips are reliable enough to create probable cause. See *supra* Section III.C.

³⁰² For example, Professor Beebe's analysis suggest that this is sometimes the case. See *supra* notes 247–250 and accompanying text.

³⁰³ For example, some commentators have argued that the Supreme Court's decision to allow courts of appeals to find that a right was not clearly established without deciding whether a violation of the constitution occurred has led to decrease in the development of constitutional law. See Hughes, *supra* note 257, at 428 & n.121; Joanna C. Schwartz, *The Case Against Qualified Immunity*, 95 NOTRE DAME L. REV. 1797, 1827 (2018) (“Indeed, courts are far more likely to grant qualified immunity motions without ruling on the underlying constitutional claim—a practice that increases constitutional stagnation, not innovation.”). A move to balancing tests might have a similar effect to preventing courts from skipping steps. By forcing courts to address all factors they might develop jurisprudence on parts of the analysis they might otherwise skip over. *But see* Beebe, *supra* note 247, at 1586–87 (finding that judges still gloss over certain factors when employing balancing tests for trademark infringement).

To some extent, the evaluation of this aspect of step tests will turn on the degree to which one agrees with the prior arguments about their flattening of normative issues. If step tests separate issues that belong together, then the development of the issue framed in the first step over the one relegated to the second may be troubling. The answer to this concern is unlikely to be global; in some cases, the weight will be appropriately placed on one factor while the others are limited in their importance as stand-alone issues. But in cases where the nature of the steps is more fraught, the balance may shift.

Further, in cases where a more rigid form of uniformity is undesirable, we may prefer to abandon step tests. While the multistep approach may lead to a type of formal uniformity, it does so at the risk of preventing lower court judges from acting to fit the law to conditions on the ground. Indeed, in some cases, this tailoring might lead to greater substantive uniformity as the law incorporates and fits local judgments to produce the desired outcome. Whether this is desirable is a case-by-case and doctrine-by-doctrine decision. Still, inhibition of pluralism in favor of rigid uniformity *can* be a cost of multistep approaches.

Above, control of lower courts was cited as a benefit of steps, and I do not wish to abandon that claim. Rather, my goal is to highlight that this benefit comes with a cost. The simplification, and therefore control, offered by steps may also come at the expense of the development offered by lower courts. Steps can channel inquiry into one area—limiting development of others—and also focus intellectual firepower at one issue among many. Thus, steps are no panacea to the problem of judicial discretion.

* * *

This cost–benefit analysis is admittedly difficult and critically context dependent. As a result, the question of stepification’s overall merit is one of judgment, rather than some ironclad conclusion one way or the other. Nevertheless, two central lessons follow from the preceding analysis. The first is that, as a broad matter, we should be more suspicious of stepification than perhaps some legal actors are. While the move towards more rigid inquiries may seem to advance the degree to which our law is *lawlike*, it also may obscure the underlying values the law is attempting to capture. This shift might be justified based on courts’ institutional competencies, and indeed the litigants in the partisan gerrymandering cases argued just that.³⁰⁴ But we

³⁰⁴ See *supra* Section III.D.

should not immediately equate workability with the notion that a given inquiry captures the values the law is intending to advance.³⁰⁵

Second, stepification should be thought of as a transsubstantive issue in jurisprudence that explains the connection between various doctrinal difficulties. While there are certainly limits to a transsubstantive approach, to the extent that stepification is a trend within American law, that trend deserves interrogation. We should not be so quick to equate a trend that seems like a natural evolution for the law with the values that we necessarily want the law to embody. Stepification is a valuable legal tool, but it, like other doctrinal technologies, comes with certain drawbacks. And indeed, in some cases these drawbacks are fatal to step tests over time, leading to their modification and potential collapse. This is the issue to which the next Part turns.

V. THE LIFE AND DEATH OF STEPIFICATIONS

This last Part aims to accomplish two goals. The first is to offer a hypothesis about the death of stepified regimes. Much of this Article has been dedicated to chronicling the rise of stepification and why it replaced older analytical modes. In light of the factors discussed in Part IV, one must ask when the faults of stepification outweigh its benefits. Having addressed this question, the second Section then attempts to sketch out alternatives to stepification. This endeavor is even more tentative than the first as it is difficult to break out of established modes of judicial lawmaking. Still, I aim to make the first moves towards other models that could replace multistep tests when they break down.

A. *The Life Cycle of Step Tests*

My normative critiques of stepification and the case studies suggest a tentative further conclusion: there is a life cycle of stepification. While some implicit stepification is necessary for the law to operate,³⁰⁶ explicit articulations like *Lemon* or *Chevron* that operate in highly contested areas of the law are often fragile and prone to tinkering or collapse as a result of their form. The aim of this final Part is to offer a brief theory of the deterioration of stepifications over time.

The main observation is that where stepification operates in normatively fraught areas, the tests that it produces are under constant strain.³⁰⁷ This is because such tests often simplify the deep normative content

³⁰⁵ Or, if we should, we should more explicitly make this argument.

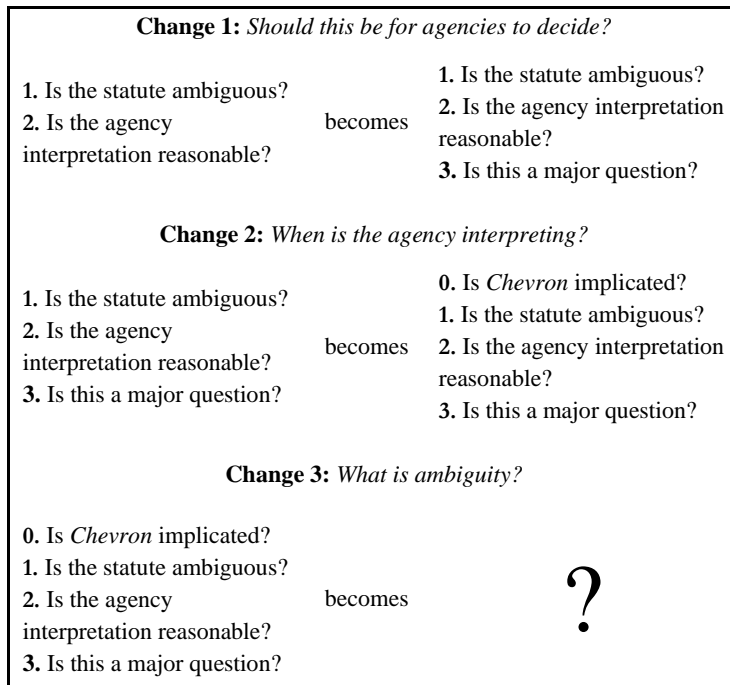
³⁰⁶ See *supra* Section I.B.

³⁰⁷ Schauer, *supra* note 36, at 804–05.

of the law.³⁰⁸ In particular, in hard cases, the factors that multistep tests assume to be separate bleed into each other, and therefore a straightforward application of the test will lead to results that seem at odds with what the law requires. This leads to the pressure to modify the test to account for these difficulties.

Here, the *Chevron* example is again instructive. While *Chevron* works well for easy cases, in harder ones it is markedly less successful, and the Court has slowly abrogated it over time.³⁰⁹ This is precisely because, in certain cases, the Court has been confronted by issues for which the framework is poorly suited. Where these cases are particularly fraught, step tests encounter difficulties and can break down over time. This process is depicted in Figure 5:

FIGURE 5: EVOLUTION OF *CHEVRON* OVER TIME³¹⁰



³⁰⁸ See *supra* Section IV.A.

³⁰⁹ See *supra* Section III.A.

³¹⁰ Change 1 represents *FDA v. Brown & Williamson Tobacco Corp.*, 529 U.S. 120, 159–61 (2000). Change 2 represents *United States v. Mead Corp.*, 533 U.S. 218, 226–27 (2001), and *Barnhart v. Walton*, 535 U.S. 212, 222 (2002). Change 3 is a hypothetical future case.

As Figure 5 shows, as the test has encountered difficult cases over time, the Court has had to adjust it by adding more and more steps and modifications to accommodate the underlying form of what the Court thinks the law *should* be.³¹¹ I would argue that this is what happened to *Auer* as a majority of the Court added further steps to constrain a doctrine that some thought had transferred too much power to agencies.³¹² In particular, the evolution of *Chevron* seems to suggest that a majority of the Court believes that a literal reading of the doctrine would apply it to too many agency actions, requiring the Court to add new steps that tell lower courts not to apply *Chevron* to both minor agency actions³¹³ and major ones.³¹⁴ This is not necessarily a bad thing, as we want the law to adjust over time. It does suggest, however, that stepification is an imperfect way to conclusively resolve issues even in cases in which the steps are relatively clear.³¹⁵ Stepification may not conclusively resolve issues, but it may provide a useful vehicle for effecting the necessary changes.

A similar, though less dramatic, tendency can be observed in the qualified immunity cases. Over time, the Court deviated from the mandatory nature of the two-step inquiry in response to critiques of that regime.³¹⁶ The excessive rigidity of the regime led to its abandonment, and though *Pearson* nominally left the two steps in place, commentators have observed that courts have in practice often focused on only one of the two (perhaps to the detriment of the development of constitutional law).³¹⁷ And yet the doctrine is still attacked by commentators who accuse it of policy failings³¹⁸ as well as a lack of fidelity to the common law in which it originated.³¹⁹ The Court may therefore be pressed to once again reconsider the two-step test,

³¹¹ People certainly can find reasons to disagree with any one of these decisions or doctrinal moves. In each case, it was a majority of the Court that thought an adjustment was needed to constrain the result of a straightforward application of the previous doctrine.

³¹² See *supra* notes 1–6 and accompanying text.

³¹³ See *supra* notes 172–176 and accompanying text.

³¹⁴ See *supra* notes 177–179 and accompanying text.

³¹⁵ This is in contrast to the *Miranda* example above where the addition of steps was in response to the need to elaborate the decision’s vague terms. See *supra* notes 278–284 and accompanying text.

³¹⁶ *Pearson v. Callahan*, 555 U.S. 223, 234 (2009) (“Lower court judges, who have had the task of applying the *Saucier* rule on a regular basis for the past eight years, have not been reticent in their criticism of *Saucier*’s ‘rigid order of battle.’” (citation omitted)).

³¹⁷ Hughes, *supra* note 257, at 428 & n.121.

³¹⁸ See, e.g., Schwartz, *supra* note 303, 1798–1800 (“If the Court did find an appropriate case to reconsider qualified immunity, and took seriously available evidence about qualified immunity’s historical precedents and current operation, the Court could not justify the continued existence of the doctrine in its current form.”).

³¹⁹ See William Baude, *Is Qualified Immunity Unlawful?*, 106 CALIF. L. REV. 45, 45 (2018).

potentially jettisoning it or adding additional steps to attempt to liberalize the doctrine while still preserving it as a shield for particular officials.

Note, it is unclear whether multistep tests actually occur more in normatively fraught areas. While high-profile tests take this form, I am unable to conclude that this is the case. Certainly, *Chevron*, qualified immunity, and strict scrutiny (and the tiers of scrutiny more generally) are important doctrinal frameworks that have taken this form. I hypothesize that it is normatively fraught areas where appellate courts value the enhanced control over lower courts provided by these frameworks to bring order to issues in which judges may be particularly tempted to reach results in line with their political priors.³²⁰

Thus, a more difficult question is whether multistep tests are more prevalent in normatively fraught areas of the law. As the data in the Appendix shows within a sample of cases, the majority of multistep tests are not found in highly contested areas. At the same time, such a simple reading of the analysis would miss the point. It may be the case that a greater *proportion* of normatively fraught areas are stepified. While an empirical confirmation of this point would be difficult, it does seem to be the case anecdotally. The application of the tiers of scrutiny, *Chevron* (and its relatives), and qualified immunity all continue to be important issues. It may be that multistep tests are particularly attractive for appellate courts in these areas because of the simplification of appellate review and the more structured analysis these tests offer to lower courts.

Of course, the fact that the Supreme Court (and many circuit courts) both decides cases and articulates doctrinal tests makes it necessary to think carefully about how to provide optimal guidance to lower tribunals in our multilevel system.³²¹ Justice Scalia advocated that higher courts should articulate holdings in a way that was more rule-like, rather than conferring discretion on lower courts by crafting narrow holdings that offered minimal guidance for future cases.³²² Stepification can therefore be seen as an attempt by judges to introduce some of the rule-like qualities that Justice Scalia advocated for in cases where the law does not admit the articulation of a true

³²⁰ See *supra* Section IV.A.

³²¹ See Scalia, *supra* note 128, at 1177 (“[T]he modern reality, at least, is that when the Supreme Court of the federal system, or of one of the state systems, decides a case, not merely the *outcome* of that decision, but the *mode of analysis* that it applies will thereafter be followed by the lower courts within that system, and even by that supreme court itself.”).

³²² *Id.* at 1178. Justice Scalia emphasized that his approach here was partially driven by the fact that the Supreme Court (and many state supreme courts) only review a fraction of lower court decisions. *Id.*

bright-line holding.³²³ But then, as the multistep test faces real-world complications, judges must attempt to rescue its rule-like quality by adding ever more innovations.

My broader hypothesis is that such systems, when put under enough pressure, eventually collapse; we are likely to see some of these systems collapse in the near future. One notable example of such a collapse occurred at the state level in the Oregon Supreme Court. In 1993, the Oregon Supreme Court announced a new three-step framework for all statutory interpretation in the case of *Portland General Electric Co. v. Bureau of Labor & Industries (PGE)*.³²⁴ The *PGE* case put forward a framework with ordered steps to evaluate the meaning of a statute.³²⁵ The courts in Oregon then followed the method as a matter of *stare decisis* for sixteen years. The regime, however, eventually gave way when the legislature acted to displace the regime by allowing courts to consider legislative history at the first step of the analysis.³²⁶ In reconsidering the stepified test, the court pointed to the views of the sponsor of the bill, who had argued that the old methodology was too harsh.³²⁷ Rephrased in the language of this Article: the steps of *PGE* were a poor fit for what the legislature thought the underlying nature of the normative issue was, and it therefore acted to re-norm the issue. Notably, the legislature chose not to reconsider or reinforce the steps of the test, but to discard steps altogether. The court interpreted the legislature as rejecting the

³²³ Cf. *id.* at 1187 (acknowledging that the “totality of the circumstances tests and balancing modes of analysis [are] with us forever” but encouraging “that those modes of analysis be avoided where possible; that the *Rule of Law*, the law of *rules*, be extended as far as the nature of the question allows”). As far as I can tell, Justice Scalia never transformed his skepticism of balancing into an endorsement of multistep tests as such, but his language here suggests that he would have frequently preferred the additional structure they offered compared to the abstract balancing of factors.

³²⁴ 859 P.2d 1143, 1145–47 (Or. 1993), *superseded by statute*, 2001 Or. Laws ch. 438, *as recognized in* *State v. Gaines*, 206 P.3d 1042 (Or. 2009) (en banc). For a discussion of the case and its role in interpretation in the state, see Abbe R. Gluck, *The States as Laboratories of Statutory Interpretation: Methodological Consensus and the New Modified Textualism*, 119 *YALE L.J.* 1750, 1775–85 (2010).

³²⁵ Roughly speaking, the framework looked at the text of the statute with respect to ordinary meaning and within the context of the statute and other provisions both within the statute and in related statutes. *PGE*, 859 P.2d at 1146. “If, but only if, the intent of the legislature is not clear from the text and context inquiry, the court will then move to the second level, which is to consider legislative history to inform the court’s inquiry into legislative intent.” *Id.* “If, after consideration of text, context, and legislative history, the intent of the legislature remains unclear, then the court may resort to general maxims of statutory construction to aid in resolving the remaining uncertainty.” *Id.*

³²⁶ See *Gaines*, 206 P.3d at 1047–48.

³²⁷ *Id.* at 1048. For another critique of the case, see Robert M. Wilsey, Comment, *Paltry, General & Eclectic: Why the Oregon Supreme Court Should Scrap PGE v. Bureau of Labor & Industries*, 44 *WILLAMETTE L. REV.* 615, 663 (2008) (“[T]he *PGE* paradigm, for all its pretense to regularity, has in fact become a distraction from the court’s work of construing statutes.”).

notion that there could be a strict hierarchy of sources in favor of a view that raised all sources to the same level.³²⁸

Likewise, one might worry that other multistep frameworks will eventually collapse under the weight of their own particular infirmities. This is especially so in cases in which the regulated actors have significant incentives to push against the boundaries of what is allowed by the test. Restated, the hypothesis could be that where hard cases are more likely to arise, stepified regimes are more likely to buckle under a weight they cannot bear.

This phenomenon mirrors what Professors Jeremy Kessler and David Pozen have observed in the case of legal theories. Professors Kessler and Pozen argue that legal theories gradually “work themselves impure” over time.³²⁹ They argue that because the theory proves unable to actually secure the values that spurred its creation in the first place, it must be modified to “fix” these issues.³³⁰ This process of refinement gradually becomes self-defeating.

The same phenomenon is observed with stepifications. They arise to manage a particular process and are sometimes advanced as a solution to the chaos bred by the previous regime.³³¹ As Professor Frederick Schauer argues when discussing the conversion of standards into rules, the imposition of rigidity into decisions may be perceived as a way to improve decisions and rescue them from the problems created when confronted with too many degrees of freedom.³³² But breakdown ensues when the proposed formalization cannot accurately capture the issue in question, similar to what Professors Kessler and Pozen observe for legal theories.

I do not mean to imply that this process is somehow unique to multistep tests. The balancing tests they replaced, for example, can be thought to have collapsed themselves. It is possible that stepification and the associated historical trends are just part of a larger cycle of doctrinal development and collapse. Further, multifactor balancing tests themselves can become unwieldy as more and more factors are added to them, to the point where they become meaningless or collapse to a small group of factors in practice.

At the same time, the rigidity of the multistep regime may make it more brittle. One downside of balancing tests can be the perception that they allow

³²⁸ *Gaines*, 206 P.3d at 1048.

³²⁹ Kessler & Pozen, *supra* note 37, at 1838.

³³⁰ *Id.* at 1839 (“Through an iterative process of contestation and reformulation, the theories become increasingly unmoored from the goals that were articulated to justify their adoption, adrift from their *raison d’être*.”).

³³¹ See *supra* Section II.D.

³³² Schauer, *supra* note 36, at 811–13.

the judge to eventually just reach their desired normative conclusion via manipulation of the relevant factors. But it is this feature that potentially makes these tests more robust to challenges. That is, the ability to warp the factors at the margin allows the judge to resolve hard cases in a way that is “just” but still “fits” with the test. To the extent multistep tests box judges in rhetorically, the end result may be for the test to collapse rather than for judges to continually reach results they feel are unjust.

This is not a one-size-fits-all theory. There are real benefits to stepification, and where those benefits significantly outweigh the harms, stepification need not break down over time. For example, in the gerrymandering context, a multiprong test such as the one advanced by the appellees in *Common Cause v. Rucho* and *Benisek v. Lamone* may represent the best way to adjudicate *most* of the claims that we care about without excessive entanglement of the courts and politics (even if that best way is ultimately not good enough).³³³ To the extent this is the case, stepifications may persist over time as second-best solutions to a particular issue that invite discontent at the margins without movement to depart from the test in general.

In instances in which stepification is merely a device to codify a normative intuition and to make cases easier, its benefits will be less clear. Where the stakes of hard cases are low, there will likely be little issue, though specific parties themselves may feel aggrieved by particular losses. Still, there will be little pressure to change the regime to deal with these minor cases. Where the hard cases matter more, there will be pressure to add additional steps to the theory over time, slowly limiting its domain or carving it apart with additional rules or exceptions.

This is not an indictment of this form of legal decision-making. Rather, the observation that these regimes tend to erode over time illustrates the reality of their costs. The issues are real and are considered by courts over time. They often lead to the abandonment of stepification. In this context, the transformation observed earlier is not an end point but rather part of a life cycle of jurisprudence that gradually mutates.

B. After Steps?

The question then is not whether multistep tests should be discarded, but how to more thoughtfully deploy and craft them to better fit the subject areas they address.

One option would be to add a catchall step in certain tests to better classify issues that arise at the boundary of the test. For example, those who

³³³ See *supra* Section III.D.

believe that the *Lemon* test is ill-suited to handle long-standing monuments could add an additional step that counsels the court to disregard the results of the inquiry if the question is close and the result would require ending a long-standing practice. Similarly, the Supreme Court could modify the qualified immunity inquiry such that if the plaintiff is able to nearly make the showing that the law was “clearly established” but the constitutional violation was significant enough, that the court could override the results of the test.

Both these proposals allow courts an off-ramp in a case in which the test is narrowly met or failed but the result may conflict with the normative goals of the doctrine. Notably, both proposals allow the court to keep the general doctrine intact for most cases while providing an escape valve for cases the doctrine gets wrong. The general test could then still help make easy cases easy without inadvertently resolving too many hard cases in the wrong way. Appellate courts could potentially caution lower courts against using these tools to override the results of ordinary cases and reserve their use for special cases that are appealed.³³⁴

Adding this additional level of flexibility might allow courts to eliminate the worst issues of multistep rigidity without disrupting its main benefits. This shift would require judges to acknowledge that some cases cannot be resolved via a neat doctrinal framework and that exceptions to the rule do exist. This change might undercut some of the legitimacy benefits that stepification confers. Still, the change may be worth it to limit other contortions that such hard cases might require.

In general, multistep tests are not always normatively fraught and may have important benefits for courts looking to create uniformity in a doctrinal area. It may further be the case that some process of doctrinal collapse and renewal will always occur in those areas of the law with the highest stakes and the toughest cases. The question should not be what will come after multistep tests but where to deploy them to achieve maximum effect and how to craft them such that their drawbacks are limited. This Section aims to provide one sketch of how to address some of those issues, but I am sure that there is further work to be done to craft even better improvements.

CONCLUSION

The central aim of this Article has been to bring the historical and normative dimensions of stepification into view. Once stepification is

³³⁴ This could, in essence, allow the highest relevant court to withhold the power to override the result of the test to itself, dictating methods for the lower courts while itself not following those impulses. Such an approach presents other interesting questions of judicial candor and its normative implications.

understood as a transdoctrinal phenomenon, its interesting historical pedigree as well as its transsubstantive implications can be understood. This Article suggests that stepification might not be as positive as it initially appears. Rather, the use of multistep tests seems most easily justified by institutional concerns within the judiciary rather than by an appeal to the normative values of the area of law the test is aiming to summarize. The process of stepification may not be an attempt to capture the precise normative content of an issue but rather a concession to the reality that other considerations sometimes make imperfect doctrinal representations the best courts can do. As such, while something may be gained from stepification, something is also lost. Thus, the case for stepification is, at best, uneasy.

APPENDIX: SAMPLE OF MULTISTEP TESTS

This Appendix aims to provide the reader with a sense of the diversity of multistep tests that are in use by the judiciary today. To this end, I searched for and examined relevant opinions published by the federal courts between 2010 and 2014.³³⁵ There were 2,210 of these cases. I then took a random sample of 100 cases and analyzed what test was being used and the number of steps.³³⁶ The remainder of this Appendix consists of those cases and the categorizations I assigned them.

Some top-line results seem worth highlighting. Of the sample, thirteen cases out of 100 were qualified immunity cases, and three were *Chevron* cases. The only other test with three occurrences was an analysis of whether there was personal jurisdiction based on minimum contacts. Additionally, three other cases analyzed the question of personal jurisdiction using a multistep test—one related to a foreign defendant in a federal question case and two related to a state long-arm statute—these cases, however, use essentially the same test looking first to the state long-arm statute and then to whether due process prohibits the exercise of jurisdiction. Finally, six tests in the sample were actually dictated by statute or regulation rather than by judges. While the number is not large enough to lead me to conclude that stepification is illusory, it does suggest that in some cases it has been dictated by other sources and that the numbers described above should not be taken literally, but rather as suggestive of the general volume of multistep tests.

³³⁵ As usual, the search term is “opinion((TWO or THREE) /1 (STEP or PRONG!) /1 (TEST or INQUIRY)).” This sample was taken on March 5, 2020. As of July 24, 2021, Lexis contains 2,212 qualifying cases from this time period, almost identical to when the sample was taken.

³³⁶ One case’s publication status was ambiguous, so it was randomly replaced. This suggests that some of the counts above may be a little overinclusive for reasons other than those discussed earlier, but I see no reason to believe the top-line results are affected as the results are again intended to be suggestive. In particular, the couple cases I was able to find with this error were tax court cases; Lexis seems to be able to perfectly sort district court and appellate court cases.

NORTHWESTERN UNIVERSITY LAW REVIEW

TABLE A1: RANDOM SAMPLE OF 100 CASES

Case	Steps	Short Description	Type
N.J. Carpenters Pension Fund v. Hous. Auth., 68 F. Supp. 3d 545, 553–54 (D.N.J. 2014)	2	Withdrawal from ERISA pension plan	Judicial
Nev. Prop. 1 LLC v. D’Amico (<i>In re</i> D’Amico), 509 B.R. 550, 556–57 (S.D. Tex. 2014)	2	Willful and malicious injury by bankruptcy debtor	Judicial
Foley v. Kiely, 602 F.3d 28, 31–32 (1st Cir. 2010)	2	Was a stop an arrest?	Judicial
United States v. Mendoza-Trujillo, 46 F. Supp. 3d 1204, 1223 (D. Utah 2014)	2	Was a stop an arrest?	Judicial
United States v. Llamas, 599 F.3d 381, 387–88 (4th Cir. 2010)	2	Vulnerable victim	Statutory or regulatory
Sheet Metal Workers Int’l Ass’n Local Union No. 27 v. E.P. Donnelly, Inc., 737 F.3d 879, 888, 891–92 (3d Cir. 2013)	3 and 2	Violation of labor law and unfair labor practice ³³⁷	Judicial
United States v. Thomas, 690 F.3d 358, 369 (5th Cir. 2012)	2	Venue	Judicial
FloorPro, Inc. v. United States, 98 Fed. Cl. 144, 147–48 (2011), <i>vacated</i> , 680 F.3d 1377 (Fed. Cir. 2012)	2	Third-party beneficiary to government contract	Judicial
Red River Res. Inc. v. Wickford, Inc., 443 B.R. 74, 82 (E.D. Tex. 2010)	2	Temporary cessation of production	Judicial
Dorchen/Martin Assocs. v. Brook of Cheboygan, Inc., 838 F. Supp. 2d 607, 611 (E.D. Mich. 2012)	2	Substantial similarity in copyright claim	Judicial
Garden Meadow, Inc. v. Smart Solar, Inc., 24 F. Supp. 3d 1201, 1209 (M.D. Fla. 2014)	2	Substantial similarity in copyright claim	Judicial
United States v. Rausch, 746 F. Supp. 2d 1192, 1199–1200 (D. Colo. 2010)	2	Substantial question in appeal	Judicial
DCFS USA, LLC v. District of Columbia, 803 F. Supp. 2d 29, 43 (D.D.C. 2011)	2 and 2	Stating a claim for municipal liability and due process	Judicial
Smith v. Fenty, 684 F. Supp. 2d 64, 67 (D.D.C. 2010)	2	Stating a claim for municipal liability	Judicial
Océ N. Am., Inc. v. MCS Servs., 748 F. Supp. 2d 481, 488–89 (D. Md. 2010)	2	State copyright claim preemption	Judicial
Ndaba v. Obama, 697 F. Supp. 2d 75, 77 (D.D.C. 2010)	3	Standing	Judicial

³³⁷ This was one of several cases in which the court applied two separate multistep tests. It first applied a three-step test to determine whether § 8(b)(4)(D) of the National Labor Relations Act had been violated and then applied a two-step test to determine whether maintaining a lawsuit constituted an unfair labor practice.

Case	Steps	Short Description	Type
Marblegate Asset Mgmt. v. Educ. Mgmt. Corp., 75 F. Supp. 3d 592, 603–04 (S.D.N.Y. 2014)	2	Standard for injunction	Judicial
Myers v. Colvin, 954 F. Supp. 2d 1163, 1169 (W.D. Wash. 2013)	5	Social Security Act disability determination	Statutory or regulatory
Mahoney v. Holder, 62 F. Supp. 3d 1215, 1221 (W.D. Wash. 2014), <i>aff'd sub nom.</i> Mahoney v. Sessions, 871 F.3d 873 (9th Cir. 2017)	2	Second Amendment rights	Judicial
Koolen v. Mortg. Elec. Registration Sys., Inc., 953 F. Supp. 2d 348, 351 (D.R.I. 2013)	3	Res judicata	Judicial
Veneruso v. Mount Vernon Neighborhood Health Ctr., 933 F. Supp. 2d 613, 632 (S.D.N.Y. 2013), <i>aff'd</i> , 586 F. App'x 604 (2d Cir. 2014)	2	Removal under 28 U.S.C. § 1444(a)(2)	Judicial
Cooley v. Lincoln Elec. Co., 693 F. Supp. 2d 767, 780 (N.D. Ohio 2010)	3	Reliable and admissible diagnosis	Judicial
Rodriguez v. Countrywide Home Loans, Inc. (<i>In re Rodriguez</i>), 517 B.R. 724, 731 (Bankr. S.D. Tex. 2014)	2	Reasonable attorney's fees for out-of-town counsel	Judicial
Standard Fire Ins. Co. v. Ford Motor Co., 723 F.3d 690, 693 (6th Cir. 2013)	2	"Rational reason" not to apply Michigan law	Judicial
Wilson v. Jara, 866 F. Supp. 2d 1270, 1289, 1302 (D.N.M. 2011), <i>aff'd</i> , 512 F. App'x 841 (10th Cir. 2013)	2 and 2	Qualified immunity and exigency for warrantless entry	Judicial
Vance v. Rumsfeld, 653 F.3d 591, 606, 611 (7th Cir. 2011), <i>rev'd en banc</i> , 701 F.3d 193 (7th Cir. 2012)	2 and 2	Qualified immunity and <i>Bivens</i> claims	Judicial
Bouchard v. Whetstone, 772 F. Supp. 2d 1352, 1356–57 (D. Colo. 2011)	2	Qualified immunity	Judicial
Brown v. Benefield, 757 F. Supp. 2d 1165, 1172–73 (M.D. Ala. 2010)	3	Qualified immunity ³³⁸	Judicial
Clements–Jeffrey v. City of Springfield, 810 F. Supp. 2d 857, 868 (S.D. Ohio 2011)	2	Qualified immunity	Judicial
Evans v. City of San Diego, 913 F. Supp. 2d 986, 995–96 (S.D. Cal. 2012)	2	Qualified immunity	Judicial
Hoover v. Walsh, 682 F.3d 481, 492 (6th Cir. 2012)	2	Qualified immunity	Judicial

³³⁸ Stepification in action! The *Brown* court added a third step to the familiar qualified immunity inquiry: before considering whether defendant violated a constitutional right and whether that right was clearly established, the court asked whether the defendant was acting within his authorized discretionary authority. Perhaps the court should have framed its first step as a step zero as in *Chevron*.

NORTHWESTERN UNIVERSITY LAW REVIEW

Case	Steps	Short Description	Type
Oyarzo v. Tuolumne Fire Dist., 955 F. Supp. 2d 1038, 1082–83 (E.D. Cal. 2013), <i>aff'd in part, rev'd in part sub nom.</i> Oyarzo v. Turner, 641 F. App'x 700 (9th Cir. 2015)	2	Qualified immunity	Judicial
Rasparido v. Carlone, 770 F.3d 97, 113 (2d Cir. 2014)	2	Qualified immunity	Judicial
Rush v. City of Mansfield, 771 F. Supp. 2d 827, 835 (N.D. Ohio 2011)	2	Qualified immunity	Judicial
Santiago v. Blair, 707 F.3d 984, 989 (8th Cir. 2013)	2	Qualified immunity	Judicial
Weeks v. Hodges, 871 F. Supp. 2d 811, 826 (N.D. Ind. 2012)	2	Qualified immunity	Judicial
Whitney v. City of Milan, 677 F.3d 292, 296 (6th Cir. 2012)	2	Qualified immunity	Judicial
Campbell v. Miller, 835 F. Supp. 2d 458, 471 (S.D. Ohio 2011)	2	Procedural due process rights	Judicial
United States v. Stearn, 597 F.3d 540, 565 (3d Cir. 2010)	3	Probable cause to search for drugs	Judicial
Park Ave. Radiologists, P.C. v. Melnick (<i>In re</i> Park Ave. Radiologists, P.C.), 450 B.R. 461, 466 (Bankr. S.D.N.Y. 2011)	2	Postconfirmation subject matter jurisdiction	Judicial
United States v. Becker, 636 F.3d 402, 405–06 (8th Cir. 2011)	4	Plain error in sentencing	Judicial
United States v. Chhun, 744 F.3d 1110, 1121 (9th Cir. 2014)	3	Plain error in jury instructions	Judicial
Carpenters Dist. Council of Kan. City Pension Fund v. JNL Constr. Co., 596 F.3d 491, 494 (8th Cir. 2010)	2	Piercing the corporate veil	Judicial
Herman v. YellowPages.com, LLC., 780 F. Supp. 2d 1028, 1030–31 (S.D. Cal. 2011)	2	Personal jurisdiction: state long-arm statute	Judicial
McCarthy v. Yamaha Motor Mfg. Corp., 994 F. Supp. 2d 1318, 1323 (N.D. Ga. 2014)	2	Personal jurisdiction: state long-arm statute	Judicial
Sec. Inv. Prot. Corp. v. Bernard L. Madoff Inv. Sec., LLC, 460 B.R. 106, 116–17, 122 (Bankr. S.D.N.Y. 2011), <i>aff'd sub nom.</i> Sec. Inv. Prot. Corp. v. Bernard L. Madoff Inv. Sec. LLC (<i>In re</i> Bernard L. Madoff Inv. Sec.), 474 B.R. 76 (S.D.N.Y. 2012)	2 and 2	Personal jurisdiction: minimum contacts and enjoining foreign action	Judicial
craigslist, Inc. v. Naturemarket, Inc., 694 F. Supp. 2d 1039, 1053 (N.D. Cal. 2010)	3	Personal jurisdiction: minimum contacts	Judicial
Thomas v. Skrip, 876 F. Supp. 2d 788, 794–95 (S.D. Miss. 2012)	2	Personal jurisdiction: minimum contacts	Judicial

Case	Steps	Short Description	Type
WesternGeco L.L.C. v. Ion Geophysical Corp., 776 F. Supp. 2d 342, 352 (S.D. Tex. 2011)	3	Personal jurisdiction: minimum contacts	Judicial
Sonera Holding B.V. v. Çukurova Holding A.Ş., 750 F.3d 221, 224 (2d Cir. 2014)	2	Personal jurisdiction: foreign defendant in federal question case	Judicial
Corning Gilbert, Inc. v. United States, 37 Ct. Int'l Trade 155, 164–65 (2013)	2	Patent infringement	Judicial
N5 Techs. LLC v. Capital One N.A., 56 F. Supp. 3d 755, 760 (E.D. Va. 2014)	2	Patent infringement	Judicial
Genusa v. Asbestos Corp., 18 F. Supp. 3d 773, 784 (M.D. La. 2014)	2	Original jurisdiction: connection test	Judicial
Mercier v. United States, 114 Fed. Cl. 795, 798 (2014), <i>aff'd in part, rev'd in part</i> , 786 F.3d 971 (Fed. Cir. 2015)	2	Money-mandating law or regulation	Judicial
<i>In re Fullenkamp</i> , 477 B.R. 826, 831 (Bankr. M.D. Fla. 2011)	2	May a debtor in possession employ an attorney?	Statutory or regulatory
Loveridge v. Barlow (<i>In re Tebbs</i>), 488 B.R. 729, 732 (D. Utah 2013)	2	Is party a transferee?	Judicial
George Fam. Tr. <i>ex rel.</i> George v. United States, 97 Fed. Cl. 625, 630 (2011)	2	Inverse condemnation	Judicial
United States v. McManaman 673 F.3d 841, 846 (8th Cir. 2012)	2	Inevitable discovery	Judicial
Moore v. Beard, 42 F. Supp. 3d 624, 633 (M.D. Pa. 2014), <i>aff'd sub nom.</i> Moore v. Sec'y Pa. Dep't of Corr., 640 F. App'x 159 (3d Cir. 2016)	2	Ineffective assistance of counsel	Judicial
Rainey v. Varner, 603 F.3d 189, 197 (3d Cir. 2010)	2	Ineffective assistance of counsel	Judicial
Quao Lin Dong v. Att'y Gen. of U.S., 638 F.3d 223, 229–30 (3d Cir. 2011)	3	Immigration: corroborating evidence	Judicial
Frank Sawyer Tr. of May 1992 v. Comm'r, 107 T.C.M. (CCH) 1316, 1318–19 (T.C. 2014), <i>supplemented by</i> 107 T.C.M. (CCH) 1621	2	Fraudulent corporate transfer	Judicial
Shannahan v. IRS, 680 F. Supp. 2d 1270, 1273–74 (W.D. Wash. 2010), <i>aff'd</i> , 672 F.3d 1142 (9th Cir. 2012)	2	FOIA: summary judgment	Judicial
Pub. Emps. for Env't Resp. v. U.S. Dep't of Com., 968 F. Supp. 2d 88, 98 (D.D.C. 2013)	2	FOIA: fee waiver	Statutory or regulatory
Corbett v. Duerring, 780 F. Supp. 2d 486, 492 (S.D. W. Va. 2011)	3	First Amendment: retaliation	Judicial
Wysocki v. Crump, 838 F. Supp. 2d 763, 767 (C.D. Ill. 2011)	3	First Amendment: retaliation	Judicial

NORTHWESTERN UNIVERSITY LAW REVIEW

Case	Steps	Short Description	Type
McComish v. Bennett, 611 F.3d 510, 520 (9th Cir. 2010), <i>rev'd sub nom.</i> Ariz. Free Enter. Club's Freedom Club PAC v. Bennett, 564 U.S. 721 (2011)	2	First Amendment protection	Judicial
Garcia v. United States, 709 F. Supp. 2d 1133, 1139 (D.N.M. 2010)	2	Federal Tort Claims Act: discretionary function	Judicial
Wenzel v. Sand Canyon Corp., 841 F. Supp. 2d 463, 474 (D. Mass 2012), <i>abrogated by</i> Culhane v. Aurora Loan Servs. of Neb., 708 F.3d 282 (1st Cir. 2013)	2	Federal Rule of Civil Procedure 19	Statutory or regulatory
Mich. Self-Insurers' Sec. Fund v. DPH Holdings Corp. (<i>In re</i> DPH Holdings Corp.), 434 B.R. 77, 82 (S.D.N.Y. 2010)	2	Excusable neglect	Judicial
Ind. Fam. & Soc. Servs. Admin. v. Ind. Fam. & Soc. Servs. Admin., Saint Catherine Hosp. of Ind., LLC (<i>In re</i> Saint Catherine Hosp. of Ind., LLC), 511 B.R. 117, 123–24 (S.D. Ind. 2014), <i>rev'd sub nom.</i> Saint Catherine Hosp. of Ind., LLC v. Ind. Fam. & Soc. Servs. Admin., 800 F.3d 312 (2015)	2	Exception to preference payment	Statutory or regulatory
Doe v. Reed, 697 F.3d 1235, 1240 (9th Cir. 2012), <i>amended by</i> Chesbro v. Best Buy Stores, L.P., 705 F.3d 913 (9th Cir. 2012)	2	Exception to mootness	Judicial
Meikle v. Olsen (<i>In re</i> Olsen), 522 B.R. 294, 325 (Bankr. D. Mont.)	2	Exception for discharge of debt	Judicial
Williams v. California, 990 F. Supp. 2d 1009, 1022 (C.D. Cal. 2012), <i>aff'd</i> , 764 F.3d 1002 (9th Cir. 2014) (<i>per curiam</i>)	3	Establishment Clause	Judicial
Wright v. La. Corrugated Prods., LLC, 59 F. Supp. 3d 767, 776 (W.D. La. 2014)	2	ERISA preemption of state law related to health-benefit plan	Judicial
John Muir Health v. Cement Masons Health & Welfare Tr. Fund for N. Cal., 69 F. Supp. 3d 1010, 1014 (N.D. Cal. 2014)	2	ERISA preemption of cause of action	Judicial
Ali v. Kasprenski, 732 F. Supp. 2d 439, 446 (D. Del. 2010)	2	Eighth Amendment: use of force	Judicial
NAACP v. N. Hudson Reg'l Fire & Rescue, 665 F.3d 464, 476–77 (3d Cir. 2011)	3	Disparate impact under Title VII of the Civil Rights Act of 1964	Judicial
Goodson v. Maggi, 797 F. Supp. 2d 624, 631 (W.D. Pa. 2011)	2	Dismissal of claim	Judicial

Case	Steps	Short Description	Type
Hall v. Life Ins. Co. of N. Am., 265 F.R.D. 356, 361–62 (N.D. Ind. 2010)	2	Discovery outside the administrative record	Judicial
United States v. Bran, 950 F. Supp. 2d 863, 871 (E.D. Va. 2013)	2	Deportation of aliens with information relevant to defense	Judicial
EEOC v. Boh Bros. Constr. Co., L.L.C., 731 F.3d 444, 462 (5th Cir. 2013)	2	Defense to vicarious employment-discrimination liability	Judicial
Metro Bank v. Kessler (<i>In re Kessler</i>), 430 B.R. 155, 161 (Bankr. M.D. Pa. 2010)	2	Core jurisdiction of bankruptcy court	Judicial
Pure Wafer, Inc. v. City of Prescott, 14 F. Supp. 3d 1279, 1297–98 (D. Ariz. 2014), <i>aff'd in part, rev'd in part sub nom.</i> Pure Wafer Inc. v. Prescott, City of, 845 F.3d 943 (9th Cir. 2017)	3	Contract Clause	Judicial
L.J. <i>ex rel.</i> N.N.J. v. Sch. Bd., 850 F. Supp. 2d 1315, 1319 (S.D. Fla. 2012)	2	Content-based challenge to independent educational plan	Judicial
Fei Mei Cheng v. Att’y Gen. of the U.S., 623 F.3d 175, 185–86 (3d Cir. 2010)	2	<i>Chevron</i>	Judicial
Mylan Pharms., Inc. v. U.S. FDA, 23 F. Supp. 3d 631, 640 (N.D. W. Va. 2014), <i>rev'd</i> , 594 F. App’x 791 (4th Cir.)	2	<i>Chevron</i>	Judicial
Nat’l Mining Ass’n v. Jackson, 768 F. Supp. 2d 34, 49 (D.D.C. 2011)	2	<i>Chevron</i>	Judicial
Gernsbacher v. Campbell (<i>In re Equip. Equity Holdings, Inc.</i>), 491 B.R. 792, 848–49 (Bankr. N.D. Tex. 2013)	2	Characterization of shareholder loans	Judicial
Shaw v. Sec’y of Dep’t of Health & Hum. Servs., 91 Fed. Cl. 715, 719 (2010)	3	Causation in off-table claim for Vaccine Act	Judicial
City Line Candy & Tobacco Corp. v. Comm’r, 141 T.C. 414, 424–25 (2013), <i>aff’d</i> , 624 F. App’x 784 (2d Cir. 2015)	2	Calculating gross receipts for purposes of the small-reseller exception	Judicial
Eurand, Inc. v. Mylan Pharms., Inc. (<i>In re Cyclobenzaprine Hydrochloride Extended-Release Capsule Pat. Litig.</i>), 676 F.3d 1063, 1085 (Fed. Cir. 2012)	2	Best mode disclosure violation	Judicial

NORTHWESTERN UNIVERSITY LAW REVIEW

Case	Steps	Short Description	Type
Inventio AG v. Thyssenkrupp Elevator Ams. Corp., 5 F. Supp. 3d 665, 675 (D. Del. 2013), <i>clarified by</i> No.08-874-RGA, 2014 WL 554853 (D. Del. 2014)	2	Best mode disclosure violation	Judicial
Klestadt & Winters, LLP v. Cangelosi, 672 F.3d 809, 817 (9th Cir. 2012)	3	Appealable interlocutory orders	Judicial
United States v. Williams, 620 F.3d 483, 488 (5th Cir. 2010)	2	Admissibility of prior wrongs	Judicial
United States v. Salomon-Mendez, 992 F. Supp. 2d 340, 342 (S.D.N.Y. 2014)	2	Admissibility of identification testimony	Judicial
Johnson v. Bd. of Trs., 666 F.3d 561, 564–65 (9th Cir. 2011)	2	ADA disability	Judicial
Draper v. Martin, 664 F.3d 1110, 1113 (7th Cir. 2011)	2	Accrual date for 42 U.S.C. § 1983 claims	Judicial
Ellsworth v. Lifescape Med. Assocs., P.C. (<i>In re</i> Ellsworth), 455 B.R. 904, 914 (B.A.P. 9th Cir. 2011)	2	Abuse of discretion in bankruptcy court	Judicial
Rediger Inv. Corp. v. H Granados Commc'ns, Inc. (<i>In re</i> H Granados Commc'ns, Inc.), 503 B.R. 726, 731–32 (B.A.P. 9th Cir. 2013)	2	Abuse of discretion in bankruptcy court	Judicial