IF THE TEXT IS CLEAR—LEXICAL ORDERING IN STATUTORY INTERPRETATION

*Adam M. Samaha**

Most courts now endorse lexical ordering for statutory cases. That is, a limited set of top-tier sources, if adequately clear, are supposed to establish statutory meaning. Lower-tier sources are held in reserve for close calls. Examples include legislative history and deference to agency positions, which often are demoted into tiebreaking roles. In fact, some such hierarchy of sources is approved by working majorities at the U.S. Supreme Court and more than forty state supreme courts. Although popular today, lexically ordered interpretation has risen and fallen before. Indeed, we should pause to reconsider whether these instructions are justified and whether judges can follow them.

This Article explores the core trade-offs and implementation challenges of lexical ordering. On trade-offs, the Article spotlights decision quality, decision costs, and, less intuitively, decisiveness. Compared to aggregating all source inferences, lexical ordering threatens decision quality by sometimes throwing out useful information, but it can reduce decision costs and probably will increase the chance of a decisive judgment. Compared to flatly excluding lower-tier sources, lexical ordering probably yields higher quality decisions and decisiveness, but also higher decision costs. Whether the overall compromise seems tolerable depends on a series of debatable judgment calls. Moreover, the actual trade-offs depend on whether judges lexically order sources in their decisionmaking, not only in their opinion writing. To date, we lack evidence either way.

The Article goes on to report results from a new vignette experiment conducted with approximately one hundred appellate judges. These judges showed curiously mixed success at lexical ordering. In a trade name case, we find little evidence that judges were improperly influenced by legislative history. In an election law case, by contrast, we find evidence that judges were improp-

* Inez Milholland Professor of Civil Liberties, New York University School of Law. Thanks to participants at Columbia Law School's Legal Theory Workshop, Duke Law School's Judicial Decisionmaking Colloquium, the 2018 Legislation Roundtable, and NYU School of Law's Faculty Workshop. For comments on an earlier draft, I am grateful to Matt Adler, Barton Beebe, Jessica Bulman-Pozen, Adam Cox, Ryan Doerfler, William Eskridge, Margaret Lemos, Marin Levy, Jamie Macleod, Caleb Nelson, David Pozen, Fred Schauer, Yahli Shereshevsky, and Peter Strauss. For discussions on the experiment reported below, I thank Emiliano Catan, Scott Hemphill, Bert Huang, Anup Malani, Jeff Rachlinski, Dan Rubinfeld, Dan Simon, Stephan Tontrup, Kristen Underhill, and Tess Wilkinson-Ryan. I owe special debts to the Institute of Judicial Administration for making the experiment possible, and to Roy Germano, Empirical Research Methodologist at NYU School of Law, for his indispensable advice and data analysis. Joy-Anette Atsegbua, Lance Bowman, Olivia Clements, and Emma Moore provided research assistance. Mistakes are mine.

erly influenced by an agency's position. There is much more to learn about the patterns of judicial behavior in this field. For now, we should expect mixed judicial success at achieving the mixed advantages and disadvantages of lexical ordering. The hard trade-offs cannot be casually assumed or ignored. With that unsettling lesson, more courts might abandon lexical ordering's complex and sometimes fragile architecture—or at least maintain respect for judges who are committed to less orthodox, more extreme, and simpler methods for deciding statutory cases.

*Where the language is plain . . . the rules which are to aid doubtful meanings need no discussion.***

*And you must strike it from your minds.****

Introduction

Abandoned at the Supreme Court for decades, lexical ordering returned to the core conventions of statutory interpretation by the late twentieth century.[1] That is, today most judges are supposed to decide whether a statute is clear using a limited set of top-tier sources and, if so, apply this meaning; if the statute remains unclear, lower-tier sources may or must be considered. Lower-tier sources are held aside just in case. Examples familiar to lawyers involve plain meaning rules that demote without condemning considerations such as legislative history, deference to administrative agencies, and the rule of lenity. Some version of this analytical strut is now planted in the approved method for deciding statutory cases in nearly every court in the country. A popular judicial turn of phrase nowadays is that legislative history "is meant to clear up ambiguity, not create it,"[2] and agency positions usually are assigned a comparable tiebreaking role.[3]

Actually, the idea of lexically ordered interpretation is no younger than Blackstone's *Commentaries*. In 1765, Blackstone listed secondary considerations for statutory cases that were supposed to matter only if "words happen to be still dubious."[4] His lower-tier sources included statutes on the same subject, along with the reason and spirit of the statute at issue.[5] Lexically ordered statutory interpretation, which can take many forms, was endorsed at the Supreme Court by 1920.[6] But that approach always had competitors. Sutherland's *Statutes* countered with an apparently wide-open aggregation of valid sources. The 1943 edition of the treatise pronounced that "statutory interpretation . . . is a fact issue. Where available, the courts should never exclude relevant evidence on that issue of fact."[7] In this regard the treatise was less Blackstone and more Holmes,[8] as were many judicial opinions in the

** Caminetti v. United States, 242 U.S. 470, 485 (1917).

*** The Verdict (20th Century Fox 1982).

1 Often I will use "interpretation" loosely to include decisionmaking related to law's meaning. *See infra* note 113.

2 Milner v. Dep't of Navy, 562 U.S. 562, 574 (2011) (Kagan, J.); *see infra* subsection I.B.1.

3 *See infra* subsection I.B.2.

4 1 William Blackstone, Commentaries *60.

5 *See id.* at *59–61.

6 *See, e.g.*, Caminetti v. United States, 242 U.S. 470, 490 (1917); *cf.* United States v. Fisher, 6 U.S. (2 Cranch) 358, 386 (1805) (Marshall, C.J.) (contrasting, apparently, plain intent with situations requiring construction).

7 2 J.G. Sutherland & Frank E. Horack Jr., Statutes and Statutory Construction § 4502, at 317 (3d ed. 1943).

8 *See* Bos. Sand & Gravel Co. v. United States, 278 U.S. 41, 48 (1928) (Holmes, J.) ("It is said that when the meaning of language is plain we are not to resort to evidence in order

1940s through the 1970s.[9] Equally important, simplification through outright exclusion of interpretive sources has been recognized as an option for centuries.[10]

Blackstone did not evaluate the trade-offs and implementation challenges for lexical ordering, of course. Nor have the rest of us done much better since then, even as the idea returned to prominence by the 1990s.[11] Indeed, judges might have become all too familiar with lexical ordering as a methodological compromise.[12] One side's persistent support for using certain sources is joined with another side's persistent objection to those sources, and a compromise is announced in which the controversial sources are not repudiated, not embraced, and not reweighted to calibrate their influence. Instead, all interpreters are asked to demote the controversial sources into a lower tier of the decision tree—with the hinge to demoted sources turning on each interpreter's sense of clarity, which is supposed to develop while demoted sources are ignored.

We ought to reconsider, now, whether this complex doctrinal architecture is justified and whether judges can follow it. Without satisfying either the judge who prefers to mix together or the judge who prefers to flatly exclude sources, the lexical ordering of sources lacks a simple and principled defense under ideal conditions, thus far.[13] Adding practical concerns about

---

to raise doubts. That is rather an axiom of experience than a rule of law, and does not preclude consideration of persuasive evidence if it exists.").

9 *See, e.g.*, Train v. Colo. Pub. Interest Research Grp., Inc., 426 U.S. 1, 10 (1976) (following *American Trucking*); United States v. Am. Trucking Ass'ns, 310 U.S. 534, 544 (1940) (indicating that legislative history may be considered "however clear the words [of a statute] may appear"); N.Y. State Bankers Ass'n v. Albright, 343 N.E.2d 735, 738 (N.Y. 1975) ("Then it is often said with more pious solemnity than accuracy, that the clarity of the statute precludes inquiry into the antecedent legislative history."); Arthur W. Murphy, *Old Maxims Never Die: The "Plain-Meaning Rule" and Statutory Interpretation in the "Modern" Federal Courts*, 75 COLUM. L. REV. 1299, 1300 (1975) (reviewing federal cases and denying a return to the *Caminetti* approach); *see also* WILLIAM N. ESKRIDGE JR., ET AL., STATUTES, REGULATION, AND INTERPRETATION 434 & n.14 (2014) (suggesting a post–World War II decline of plain meaning rules in state courts, alongside a rise in use of legislative history, then a rebound for plain meaning); *id.* at 350 (suggesting that many courts stuck with plain meaning).

10 *See, e.g.*, Millar v. Taylor (1769) 98 Eng. Rep. 201, 217; 4 Burr. 2303, 2332 (Eng.) (Willes, J.) (regarding textual changes during parliamentary proceedings); Max Radin, *Statutory Interpretation*, 43 HARV. L. REV. 863, 872–73 (1930).

11 *See infra* Section I.B.

12 *See* Abbe R. Gluck, *The States as Laboratories of Statutory Interpretation: Methodological Consensus and the New Modified Textualism*, 119 YALE L.J. 1750, 1758 (2010) (describing a compromise that prioritizes textual analysis over legislative history and that is viable in state courts). Gluck's careful case studies are real contributions; the article did not aspire to specify all trade-offs or test whether judges lexically order in practice. *See id.* at 1770.

13 *See infra* Section II.A (sketching arguments in ideal decision situations). An earlier effort to analyze trade-offs appears in Adam M. Samaha, *On Law's Tiebreakers*, 77 U. CHI. L. REV. 1661, 1665–71, 1689–1700, 1708–17 (2010) [hereinafter Samaha, *Tiebreakers*], which did not confront implementation challenges. For a recent contribution with several incisive observations that are broadly consistent with *Tiebreakers*, see William Baude & Ryan D. Doerfler, *The (Not So) Plain Meaning Rule*, 84 U. CHI. L. REV. 539 (2017) (discussed in

decision costs and judicial decisiveness tends to strengthen the argument for lexical ordering, as we shall see. But the upsides still might not be worth the effort after considering the downside risk that useful information will be ignored.[14] The trade-offs should be troubling.

On the other hand, perhaps the instructions do not matter. Judges and others have suggested repeatedly that there is nothing clear about clarity tests for statutes.[15] Perhaps judges have convenient understandings of "clarity" that allow each judge to consider lower-tier sources when they want to and not when they do not. Another possibility is that judges are unable to follow the instructions regardless of their preferences. "When the reading is done and the case has been analyzed and argued," Judge Randolph asked twenty years ago, "how can it be said that the judge turned to the legislative history only after finding the statutory language ambiguous?"[16] The same question

---

subsection II.A.1). For an older critique of plain meaning rules compared to inclusion and exclusion rules, see Murphy, *supra* note 9, at 1316.

14 *See infra* Section II.B.

15 *See* Brett M. Kavanaugh, *Fixing Statutory Interpretation*, 129 HARV. L. REV. 2118, 2136, 2138 (2016) (book review) ("The simple and troubling truth is that no definitive guide exists for determining whether statutory language is clear or ambiguous."). For earlier expressions of the idea, see, for example, Exxon Mobil Corp. v. Allapattah Servs., Inc., 545 U.S. 546, 572 (2005) (Stevens, J., dissenting) ("[A]mbiguity is apparently in the eye of the beholder . . . ."); State *ex rel.* Kalal v. Circuit Court for Dane Cty., 681 N.W.2d 110, 130 (Wis. 2004) (Abrahamson, C.J., concurring) (similar); HENRY M. HART, JR. & ALBERT M. SACKS, THE LEGAL PROCESS 1237 (William N. Eskridge, Jr. & Philip P. Frickey eds., Foundation Press 1994) (1958) (asking whether the criteria for plain meaning were sufficiently definite to block the influence of result-oriented preferences); Frederick J. de Sloovère, *Textual Interpretation of Statutes*, 11 N.Y.U. L. Q. REV. 538, 548 (1934) (asserting that authorities leave no "hint as to what is the test of explicitness"); *see also* ANTONIN SCALIA & BRYAN A. GARNER, READING LAW: THE INTERPRETATION OF LEGAL TEXTS 436 (2012) ("[D]etermining what is unambiguous is eminently debatable."); James J. Brudney, *Confirmatory Legislative History*, 76 BROOK. L. REV. 901, 902 (2011) (suggesting that a reader's self-confidence can influence clarity perceptions); Thomas W. Merrill, *Step Zero After* City of Arlington, 83 FORDHAM L. REV. 753, 753 (2014) (asserting that the *Chevron* doctrine's clarity and reasonableness elements "are sufficiently flexible to permit virtually any outcome"); George H. Taylor, *Structural Textualism*, 75 B.U. L. REV. 321, 356–58, 357 n.163 (1995) (collecting sources). A distinct concern is ambiguity over which sources a plain meaning rule is supposed to include. *See* REED DICKERSON, THE INTERPRETATION AND APPLICATION OF STATUTES 229–31 (1975).

16 A. Raymond Randolph, *Dictionaries, Plain Meaning, and Context in Statutory Interpretation*, 17 HARV. J.L. & PUB. POL'Y 71, 76 (1994); *see* WILLIAM N. ESKRIDGE JR., INTERPRETING LAW 400 (2016) (maintaining that judges in fact develop plain meaning from an examination of all relevant legal sources); Abbe R. Gluck & Richard A. Posner, *Statutory Interpretation on the Bench: A Survey of Forty-Two Judges on the Federal Courts of Appeals*, 131 HARV. L. REV. 1298, 1325 (2018) ("[E]ven judges who may not seek out legislative history are always exposed to it through briefing materials, so it necessarily helps form an impression . . . ."); Ethan J. Leib & Michael Serota, Essay, *The Costs of Consensus in Statutory Construction*, 120 YALE L.J. ONLINE 47, 62 (2010) ("[T]he text in the hard cases will still be viewed by many judges through the lens of legislative history."); Adrian Vermeule, *Legislative History and the Limits of Judicial Competence: The Untold Story of* Holy Trinity Church, 50 STAN. L. REV. 1833, 1882 & n.170 (1998) (quoting Judge Randolph); *see also* HART & SACKS, *supra* note 15, at

should be asked about every lower-tier source.[17]

Before another twenty years pass, we need to make additional theoretical and empirical progress on all of the above ideas. This Article aims to help by exploring multiple aspects of lexical ordering in statutory cases: its resurgence, its unique trade-offs when followed, and its implementation challenges, alongside new experimental evidence. These matters are tightly connected. Lexical ordering's spread increases the urgency of revisiting its core trade-offs, which are shaped by its ground-level implementation. There will be much good work left over, in statutory interpretation and in other fields where lexical ordering might be adopted.[18] But the work below contributes on multiple fronts. More broadly, efforts like this support thoughtful evaluation of the intertwined architectures of interpretation—doctrinal, cognitive, institutional, physical—by taking theory and implementation seriously, and together. There can be no simple and definitive take on lexical ordering for now. We should recognize this, even advertise it, as motivation to continue rethinking the foundations of statutory interpretation within realistic decision environments.

Part I reviews the logic of lexical ordering and shows how far it has spread into the conventions for deciding statutory cases. The discussion focuses on the position of legislative history and agency interpretations. Joining working majorities at the U.S. Supreme Court, more than forty state supreme courts now endorse lexical inferiority for state-level legislative history, and approximately thirty-five do so for state agency interpretations. But official positions on interpretive method have not been stable over long periods in the past, whatever effect they have on behavior.

Part II identifies hard trade-offs, assuming that lexical ordering instructions are followed and that the clarity test is diagnostic.[19] Compared to aggregating all relevant source inferences, lexically ordering sources threatens decision quality, but it can reduce decision costs and probably will increase decisiveness. Compared to flatly excluding lower-tier sources, lexical ordering probably yields higher quality decisions and increased decisive-

1236 (contending that the plain meaning rule inserted "an artificial and curious dualism into arguments and opinions").

17 *See* KENT GREENAWALT, STATUTORY AND COMMON LAW INTERPRETATION 90 (2013) (suggesting empirical study to help measure judicial adherence to lexical ordering and the effect on lawyering); Adrian Vermeule, *Interpretive Choice*, 75 N.Y.U. L. REV. 74, 82, 84 (2000) (indicating that lexical ordering is an unavoidable methodological question for statutory cases that deserves attention in light of high theory and limited evidence); *infra* Section III.A (discussing Ward Farnsworth et al., *Ambiguity About Ambiguity: An Empirical Inquiry into Legal Interpretation*, 2 J. LEGAL ANALYSIS 257 (2010), which involved student respondents, and other studies involving judges).

18 Possible extensions range widely to include contracts, patents, treaties, and more. *See* Samaha, *Tiebreakers*, *supra* note 13, at 1681, 1700–37 (addressing securitized debt, affirmative action, statutory and constitutional interpretation, and legal institutions writ large).

19 *See infra* text accompanying notes 123–24 (discussing the probabilistic identification of useful lower-tier sources, sight unseen).

ness, but also higher decision costs. Effects on the probability of a decisive outcome, in particular, are not widely recognized today. Adding indecision risks to decision costs does generate plausible arguments for demoting sources such as agency interpretations and legislative history. Perhaps the overall compromise is tolerable to methodological moderates. But these debatable judgment calls should be recognized for what they are. Moreover, the actual trade-offs depend on whether judges lexically order sources in their decisionmaking, not only in their opinion writing. To date, we lack on-point evidence either way.

Part III turns to implementation and evidence. It presents the results of a new survey and experiment conducted with approximately one hundred appellate judges. The survey offers only limited evidence that judges' propensities to find statutory text clear are aligned with their general views about the usefulness of legislative history.[20] Instead, the survey suggests that lexical ordering entails sacrifice from many judges. In the vignette experiment, though, judges showed curiously mixed success at ignoring lower-tier sources.[21] In a trade name case, we find little evidence that judges were improperly influenced by legislative history. In an election law case, by contrast, we do find evidence that judges were improperly influenced by an agency's position. Even in the election law case, however, we cannot confirm that judges conveniently bent the clarity test to change the legal relevance of attractive or unattractive sources. Other kinds of failure are possible—including that exposure to a lower-tier source will, distractingly, boost the judge's assessment of statutory clarity with or without final judgments flipping.[22]

There is more to learn. But our results do suggest the outlines of actionable lessons. Sometimes judges will indeed succeed at lexical ordering (perhaps when the case is run-of-the-mill and the lower-tier source presents a complication) in which case the instructions will bite and the hard trade-offs of quality, cost, and decisiveness should be confronted. At other times, judges will fail (perhaps when the case is ideologically charged and the lower-tier source category is attractive) in which case lexical ordering instructions are at best wasteful distractions. As we build theory and evidence to better identify domains of success and failure, we should generally expect mixed judicial success at achieving the mixed advantages and disadvantages of lexically ordered statutory interpretation. Maybe that much knowledge is enough to unsettle current trends. More courts might sensibly abandon lexical ordering's complex and sometimes fragile architecture—or at least maintain respect for those judges who are committed to less orthodox, more extreme, and simpler methods.[23]

---

20 *See infra* Section III.B and Figure 1 (showing the dispersion of responses).

21 *See infra* subsection III.C.3, Figures 2–5, and Tables 3–4.

22 *See infra* subsection III.D.2.

23 *See infra* subsection III.D.3. The challenges for lexical ordering generally do not depend on which sources are placed in which tiers. True, each source category—precedent, canons, legislative history, agency positions, and so on—may present special chal-

I. CONCEPTS AND TRENDS

A. Lexical Ordering

Successful lexical ordering creates a particular kind of priority among considerations that might be used to make a single decision. Higher-tier (lexically superior) considerations must be analyzed on their own and must trump all lower-tier (lexically inferior) considerations, which are excluded from the analysis unless the top-tier considerations are deemed inadequate to deliver a unique result.[24] Part of the trick is determining what counts as an adequate basis for decision using only higher-tier sources.

In any event, lexical ordering is not sequencing.[25] Sequencing rules tell us when, not whether, considerations are supposed to be taken up. Lexical ordering's special contribution to decision theory is the prioritization of one set of considerations such that others might or might not be ruled out. Sequencing a list of considerations from first to last might be necessary for people to make decisions, but lexical ordering certainly is not. And imposing a sequencing rule is compatible with guaranteeing that every possible basis for decision will be considered, but lexical ordering really is not.

This might look a bit complicated, and it is, but the logic of lexical ordering is no more foreign than the rules for alphabetizing.[26] Alphabetizing requires us to compare only the first letters in two words and ignore all other letters unless the first letters are the same; the influence of subsequent letters is conditioned on earlier appearing letters yielding ties that need breaking. Gracefully and powerfully, the algorithm sorts every uniquely spelled word in the universe into its own, easily located, objectively identified ordinal position. So nobody doubts the usefulness of lexical ordering for organizing bookshelves (where books are still in use). Whether lexical ordering is good for much else, however, ought to be controversial.

Indeed, lexical ordering often is beside the point. We face long running and unavoidable debates about the legitimacy and value of various sources in judicial decisionmaking. For example, some people contend that legislative

---

lenges for judicial use. But this Article concerns the architecture of lexical ordering and is not a categorical attack on or defense of any particular type of source.

24 *See* Samaha, *Tiebreakers*, *supra* note 13, at 1669 (defining tiebreakers strictly as lexically inferior decision rules).

25 On sequencing without lexical ordering, see Adam M. Samaha, *Starting with the Text—On Sequencing Effects in Statutory Interpretation and Beyond*, 8 J. LEGAL ANALYSIS 439, 441–43 (2016) (emphasizing cognitive effects of sequencing, including conditions under which text-first sequencing might backfire). Lexical ordering can be accomplished in conjunction with any sequencing rule, but some sequencing rules will be wasteful and could make successful lexical ordering more difficult. *See id.* at 441 n.1.

26 On the early development of alphabetical arrangements, see LLOYD W. DALY, CONTRIBUTIONS TO A HISTORY OF ALPHABETIZATIONS IN ANTIQUITY AND THE MIDDLE AGES 15–26 (1967) (reviewing evidence from libraries and cult-membership lists in ancient Greece). For a famous application to normative judgment, see JOHN RAWLS, A THEORY OF JUSTICE 42–43 & n.23 (1971) (depicting the categorical normative superiority of certain considerations in evaluating political systems).

history such as committee reports and floor statements should never influence judges,[27] while others maintain that such sources should be taken for whatever they are worth within a larger mix of considerations.[28] Parallel disagreements recur over the appropriate influence of administrative agency positions.[29] Then there are the challenges of weighting and integrating the implications of whichever sources are deemed valid, in accord with whichever larger roles judges take on when deciding statutory cases. These debates are necessary, even healthy, however dated.

And none of these competing positions suggest lexical ordering. As a matter of principle and logic, interpreters can exclude illegitimate, irrelevant, or otherwise disfavored sources of information, and they can include various sources to build an aggregated mix of weighted inferences—all without lexical ordering. Lexically ordered interpretation neither fully excludes nor fully includes lower-tier sources. It swings between exclusion and inclusion depending on a judge's estimations of clarity. It therefore presents a unique blend of trade-offs that depend heavily on which sources are assigned to which tiers, the content of the clarity test for moving between tiers, and how judges implement these instructions.

Maybe the effort is worthwhile. But we can begin by noting that neither the trade-offs nor the prospects for successful implementation are anything like alphabetizing books on shelves.

B. Widespread Endorsement

Whatever the complications, we should appreciate the impressive spread of lexical ordering within officially approved methods for deciding statutory cases. By the late stages of the twentieth century, lexically ordered statutory interpretation was trendy. The discussion here concentrates on two prominent strains, involving legislative history and agency interpretations, but the idea is more widespread in statutory cases.[30] Worth underscoring is that the most recent rise of the idea has been promoted by judges, not legislatures.

---

27 *See, e.g.*, Bank One Chi., N.A. v. Midwest Bank & Tr. Co., 516 U.S. 264, 279–80, 283 (1996) (Scalia, J., concurring in part and concurring in the judgment).

28 *See, e.g.*, Exxon Mobil Corp. v Allapattah Servs., Inc., 545 U.S. 546, 572 (2005) (Stevens, J., joined by Breyer, J., dissenting).

29 *Compare* Golden Rule Ins. Co. v. Tomlinson, 335 P.3d 1178, 1188 (Kan. 2014) ("No deference is paid to an agency's statutory interpretation."), *with* Jack M. Beermann, *End the Failed* Chevron *Experiment Now: How* Chevron *Has Failed and Why It Can and Should Be Overruled*, 42 CONN. L. REV. 779, 779 (2010) (advocating a revival of pre-*Chevron* multifactor inquiry), *and* Kevin M. Stack, Essay, *An Administrative Jurisprudence: The Rule of Law in the Administrative State*, 115 COLUM. L. REV. 1985, 1998 (2015) (observing that many people view the statutory clarity question as a matter for de novo court judgment), *and* Peter L. Strauss, Essay, *"Deference" Is Too Confusing—Let's Call Them "*Chevron *Space" and "*Skidmore *Weight,"* 112 COLUM. L. REV. 1143, 1164–65 (2012) (arguing that judges should give some weight to agency judgments about the boundaries of their statutory authority).

30 For the demotion of general statutory purpose and policy-related consequences, see Sebelius v. Cloer, 133 S. Ct. 1886, 1895–96 (2013), and John F. Manning, *The New Purposivism*, 2011 SUP. CT. REV. 113, 126–29. For the further demotion of the rule of lenity to a last-

The state-level construction acts that address lexical ordering are few, and a few of those statutes have been effectively ignored in court.

1. Legislative History

Congress has not produced an overarching construction act, but federal judges largely have embraced lexical ordering for federal statutory cases. An example involves legislative history such as committee reports and floor statements. Federal judges often tell us that these sources are ignored unless the statute's meaning is otherwise unclear.[31] Thus *Matal v. Tam*[32] recently asserted that "our inquiry into the meaning of the statute's text ceases when 'the statutory language is unambiguous and the statutory scheme is coherent and consistent.'"[33] Yes, the opinion went on to conclude that the party pointing at the legislative history of the trademark statute had not come up with anything telling.[34] As well, some of the Supreme Court's modern remarks—that legislative history "need not" be considered—are less than definitive.[35] But working majorities do appear committed to the lexical inferiority of legislative history.

The history of this commitment is both long and unstable, however. Lexical ordering around a plain meaning rule gained judicial support during the mid-1800s,[36] and was featured in Supreme Court opinions by the early

---

ditch tiebreaker, see, for example, Ocasio v. United States, 136 S. Ct. 1423, 1434 n.8 (2016).

31 *See* ROBERT A. KATZMANN, JUDGING STATUTES 45 (2014) (indicating that federal court opinions now commonly declare that clear statutory language marks the end of statutory interpretation). There are many current examples of courts taking this approach. *See, e.g.*, *In re* Del Biaggio, 834 F.3d 1003, 1010 (9th Cir. 2016) (O'Scannlain, J.) ("Only when statutes are ambiguous may courts look to legislative history."); United States v. Rowland, 826 F.3d 100, 108 (2d Cir. 2016) (Carney, J.) (similar); Am. Fed'n of Gov't Emps., Local 3669 v. Shinseki, 709 F.3d 29, 35 (D.C. Cir. 2013) (Sentelle, J.) (similar). Consider also the less sweeping statements in AquAlliance v. U.S. Bureau of Reclamation, 856 F.3d 101, 105 (D.C. Cir. 2017) (Millett, J.), which suggested that it was following the Supreme Court's indication that legislative history may not be used to create ambiguity in the Freedom of Information Act, and Lee v. Norfolk S. Ry. Co., 802 F.3d 626, 633 (4th Cir. 2015) (Floyd, J.), which states that consideration of legislative history is unnecessary if statutory text is unambiguous.

32 137 S. Ct. 1744 (2017).

33 *Id.* at 1756 (quoting Barnhart v. Sigmon Coal Co., 534 U.S. 438, 450 (2002)).

34 *Id.* Proceeding to consider legislative history, even after asserting statutory clarity, is not unusual. *See, e.g.*, *AquAlliance*, 856 F.3d at 105; *Del Biaggio*, 834 F.3d at 1010–11; *Rowland*, 826 F.3d at 109; *Lee*, 802 F.3d at 633.

35 *E.g.*, Trump v. Hawaii, 138 S. Ct. 2392, 2412 (2018); State Farm Fire & Cas. Co. v. United States *ex rel.* Rigsby, 137 S. Ct. 436, 444 (2016); United States v. Woods, 134 S. Ct. 557, 567 n.5 (2013) (similar); *see also* Murphy, *supra* note 9, at 1303–04 (suggesting that such "need not" language does not support a plain meaning rule).

36 *See* William S. Blatt, *The History of Statutory Interpretation: A Study in Form and Substance*, 6 CARDOZO L. REV. 799, 812–13, 813 nn.73–74 (1985) (reviewing treatises and state and federal judicial opinions).

1900s. *Caminetti v. United States*,[37] a textbook example, declared it "elementary that the meaning of a statute must, in the first instance, be sought in the language in which the act is framed."[38] This statement looks like a sequencing rule, but the Court followed up with an endorsement of lexical priority: "[I]f that [meaning] is plain, . . . the sole function of the courts is to enforce it according to its terms," and "the rules which are to aid doubtful meanings need no discussion."[39] *Caminetti* may be contrasted with the *Holy Trinity* approach of the late 1800s,[40] because the latter emphasized the law's spirit over its text while the former did the opposite.[41] But on the propriety of lexical ordering, *Caminetti* and *Holy Trinity* are in lockstep. Their approaches differ over which sources should be placed in which tiers, not whether tiers should be constructed in the first place.

Lexical ordering was apparently replaced by an ultrainclusive "no rule of law" approach in *United States v. American Trucking Associations* in 1940.[42] Emphasizing legislative purpose and relying on legislative history, the Court's opinion announced that "[w]hen aid to construction of the meaning of words, as used in the statute, is available, there certainly can be no 'rule of law' which forbids its use, however clear the words may appear on 'superficial examination.'"[43] This add-it-all-up notion was reaffirmed by a unanimous Supreme Court as late as 1976, when a lower court was rebuked for saying exactly what the high Court would later endorse once again: that judges

37 242 U.S. 470 (1917).

38 *Id.* at 485.

39 *Id.*

40 *See* Church of the Holy Trinity v. United States, 143 U.S. 457, 459 (1892) ("[A] thing may be within the letter of the statute and yet not within the statute, because not within its spirit, nor within the intention of its makers.").

41 *See* JOHN F. MANNING & MATTHEW C. STEPHENSON, LEGISLATION AND REGULATION 49 (3d ed. 2017) (presenting *Holy Trinity*–style purposivism as competing with a *Caminetti*-style plain meaning rule); *cf.* Adrian Vermeule, *The Cycles of Statutory Interpretation*, 68 U. CHI. L. REV. 149, 183 (2001) (describing *Holy Trinity*, *Caminetti*, and *American Trucking* as part of a cycle between relatively literalist and purposivist approaches, the latter allowing many sources extrinsic to statutory text).

42 United States v. Am. Trucking Ass'ns, 310 U.S. 534, 544 (1940) (internal quotation marks omitted).

43 *Id.* at 543–44 (quoting Helvering v. N.Y. Tr. Co. 292 U.S. 455, 465 (1934)); *see id.* at 547–48 (proceeding to rely on legislative history and other sources); HART & SACKS, *supra* note 15, at 1237 (asserting "the overthrow of the plain meaning rule in the federal courts"). *But cf. Ex parte* Collett, 337 U.S. 55, 61 (1949) ("[T]here is no need to refer to the legislative history where the statutory language is clear."). For a detailed argument that the capacity and lawyering of federal agencies help explain the Court's rising reliance on legislative history, see Nicholas R. Parrillo, *Leviathan and Interpretive Revolution: The Administrative State, the Judiciary, and the Rise of Legislative History, 1890–1950*, 123 YALE L.J. 266, 315 (2013).

"need not resort to legislative history" when a statute is "plain and unambiguous."[44]

By the early 1990s, lexical inferiority for legislative history reappeared in Supreme Court opinions, as others have recounted.[45] This development should be understood alongside concentrated attacks on the intelligibility, reliability, and legitimacy of legislative history as a reflection of collective congressional intent. Those attacks influenced judicial thinking, but not enough to achieve a flat exclusion. The demotion of legislative history into a lexically inferior tier is now endorsed by many federal judges who are not self-described textualists. "[W]e do not resort to legislative history to cloud a statutory text that is clear," Justice Ginsburg wrote for the Court in 1994.[46] When Congress bothers to produce legislative history,[47] the Supreme Court tends to demote it without necessarily excluding it.

State courts are where most judicial interpretation takes place in this country.[48] While their approved methods vary, more than forty state supreme courts more or less plainly announce that they push legislative history into a lower tier.[49] Thus Michigan's Supreme Court has announced that, when statutory text is "unambiguous . . . the examination of legislative history 'of any form' is not proper."[50] Sometimes a state court's position is not so categorical, as when California's high court recently called a statute

44 Colo. Pub. Interest Research Grp., Inc. v. Train, 507 F.2d 743, 748 (10th Cir. 1974), *rev'd*, 426 U.S. 1, 10–24 (1976) (quoting *American Trucking* and relying heavily on a committee report and floor debates).

45 *See, e.g.*, Manning, *supra* note 30, at 126; Vermeule, *supra* note 41, at 183; *see also* W. Va. Univ. Hosps., Inc. v. Casey, 499 U.S. 83, 98–99 (1991). Compare less categorical statements in the 1980s and late 1970s. *See, e.g.*, United States v. Ron Pair Enters., Inc., 489 U.S. 235, 242 (1989) (indicating that a statute's plain meaning should be conclusive "except" if the result would be "demonstrably at odds with the intentions of its drafters" (quoting Griffin v. Oceanic Contractors, Inc. 458 U.S. 564, 571 (1982))); Tenn. Valley Auth. v. Hill, 437 U.S. 153, 184 n.29 (1978) (stating that "we ordinarily" ignore legislative history when a statute is facially unambiguous).

46 Ratzlaf v. United States, 510 U.S. 135, 147–48 (1994).

47 *See* BARBARA SINCLAIR, UNORTHODOX LAWMAKING 18, 51 (5th ed. 2017) (reporting that major measures bypassed committees in the House about one-third of the time from 2007 to 2014, and in the Senate more than half the time from 2009 to 2014).

48 *See* ESKRIDGE, *supra* note 16, at 199–200, 244–45 (reviewing availability of and reliance on state-level legislative history).

49 Thirty-seven state supreme courts plainly adopt this position, in my view; approximately eight other supreme courts are somewhat harder to code but apparently support such lexical ordering. These observations are based on a nationwide review of state construction acts and state supreme court opinions from 2013 to 2017, plus several earlier opinions where necessary or helpful to characterize a state's most recent position, and from 1973 to 1977. A database of authorities with Westlaw search strings is on file with the author.

50 Aroma Wines & Equip., Inc. v. Columbian Distrib. Servs., Inc., 871 N.W.2d 136, 146 n.49 (Mich. 2015) (quoting *In re* Certified Questions from the U.S. Court of Appeals for the Sixth Circuit, 659 N.W.2d 597, 601 n.5 (Mich. 2003)) (discussing legislative staff analysis); *cf.* ESKRIDGE, *supra* note 16, at 202 (observing a march "toward the exclusionary approach since 1999" in Michigan).

clear and said "we need go no further," then proceeded to check on relevant legislative history.[51] Nonetheless, most state courts now align with the Supreme Court on this issue. In the mid-1970s, when the *American Trucking* approach was reaffirmed for federal courts, explicit state court commitments to the lexical inferiority of legislative history did appear but were less dominant. In those years, many state supreme courts were not articulating a position on the matter.[52]

Today, only a few state courts are bold enough to explicitly reject lexical inferiority for legislative history. New York remains outside the trend, for instance, although the state's position has not been fixed for long periods. Recently the high court affirmed that, "[a]lthough the plain language of the statute provides the best evidence of legislative intent, 'the legislative history of an enactment may also be relevant and is not to be ignored, even if words be clear.'"[53] Alaska's judiciary is a resolute outlier on this matter, having articulated a "sliding scale approach" to plain meaning in conjunction with other sources,[54] while New Mexico's high court has acknowledged extraordinary situations in which plain meaning "must yield" to indications from legislative history.[55] But in the past few years, only a couple of other state supreme courts have expressed even limited support for legislative history entering the top tier.[56]

At the opposite end of the spectrum, the number of state judiciaries that flatly exclude legislative history appears to be zero. On occasion, however, the stakes are low. "Legislative history" can be a very limited set of material at

51 Scher v. Burke, 395 P.3d 680, 688 (Cal. 2017) (quoting Microsoft Corp. v. Franchise Tax Bd., 139 P.3d 1169, 1173 (Cal. 2006)) (proceeding to rely on a legislative counsel's digest, the enrolled bill memorandum to the governor, and committee analysis).

52 Authorities are gathered in the file referenced in note 49.

53 Kimmel v. State, 80 N.E.3d 370, 377 (N.Y. 2017) (quoting Tompkins Cty. Support Collection Unit v. Chamberlin, 786 N.E.2d 14 (N.Y. 2003)). Compare the less committed position in Avella v. City of New York, 80 N.E.3d 982, 989 (N.Y. 2017), in which the court indicated that the court "need not" consider legislative history because of the statute's plain language, but proceeded to consider it anyway, and contrast the earlier declaration in Wash. Post Co. v. N.Y. State Ins. Dep't, 463 N.E.2d 604, 606 (N.Y. 1984), in which the court said: "When the plain language of the statute is precise and unambiguous, it is determinative." *Id.*

54 Muller v. BP Expl. (Alaska), Inc., 923 P.2d 783, 788–89 (Alaska 1996); *see* Alaska Miners Ass'n v. Holman, 397 P.3d 312, 315 (Alaska 2017) (indicating that the court considers legislative history alongside statutory text).

55 Fowler v. Vista Care, 329 P.3d 630, 633–34 (N.M. 2014) (quoting Sims v. Sims, 930 P.2d 153, 157 (N.M. 1996)) (referencing legislative history, equity, and other sources).

56 Oregon might have turned against lexical ordering, too. *See infra* note 69. For tension in recent pronouncements within the same state, compare, for example, AIDS Support Grp. of Cape Cod, Inc. v. Town of Barnstable, 76 N.E.3d 969, 974–75 (Mass. 2017) (quoting Hoffman v. Howmedica, Inc., 364 N.E.2d 1215, 1218 (Mass. 1977)), which stated that legislative history is "not ordinarily" a proper source of construction if statutory language is unambiguous, and consulted such sources, with People for the Ethical Treatment of Animals, Inc. v. Dep't of Agric. Res., 76 N.E.3d 227, 234–35 (Mass. 2017), which stated that clear statutory language is conclusive.

the state level, perhaps referring to earlier enacted versions of the state code and nothing else. A few states do not maintain official legislative history in the sense of witness testimony, legislator debates, committee reports, or staff summaries.[57] Yet lexical ordering is the prevailing approach in most states nonetheless. That is, a few states demote earlier enacted versions of statutes into a lower tier, having no other official legislative history to demote.[58]

Regardless, judges rather than legislatures lead the way. Few state codes of construction address whether legislative history should be lexically inferior to other sources.[59] The state construction acts that do speak to the role of legislative history are a bit mixed. The Texas code provision, adopted in 1985, looks like support for legislative history staying in the top tier: "In construing a statute, whether or not the statute is considered ambiguous on its face, a court may consider among other matters the . . . legislative history."[60] The Oregon code, as amended in 2001, backs the ability of parties to offer legislative history but defers to judges on how much weight to give it.[61] In contrast, the Pennsylvania code, adopted in 1972, obviously embraces lexical ordering. The code announces that when a statute's words are "clear and free from all ambiguity, the letter of it is not to be disregarded under the

57 *See, e.g.*, State v. $223,405.86, 203 So.3d 816, 832 n.8 (Ala. 2016) ("Alabama does not create or maintain typical legislative-history material such as committee reports and records of hearings."); Caves v. Yarbrough, 991 So.2d 142, 153 n.17 (Miss. 2008) (similar); Heath v. Guardian Interlock Network, Inc., 369 P.3d 374, 378 (Okla. 2016) (similar); State *ex rel.* Biafore v. Tomblin, 782 S.E.2d 223, 234 (W. Va. 2016) (Loughry, J., concurring) (similar).

58 *See, e.g.*, *Ex parte* B.C., 178 So.3d 853, 855–56 (Ala. 2015) (demoting "the law as it existed prior to such statute's enactment" (quoting Reeder v. State *ex rel.* Myers, 314 So.2d 853, 857 (1975))). *But cf.* Ray v. Swager, 903 N.W.2d 366, 380–82 & nn.68 & 74 (Mich. 2017) (distinguishing legislative history from statutory history and allowing statutory amendment history into the top tier); Columbia Riverkeeper v. Port of Vancouver USA, 395 P.3d 1031, 1038 (Wash. 2017) (including statutory amendments in the top tier).

59 On eleven state construction acts that do address legislative history, see Jacob Scott, *Codified Canons and the Common Law of Interpretation*, 98 GEO. L.J. 341, 380–82 (2010). Legislative silence on this point is not entirely for lack of trying. The American Bar Association's National Conference of Commissioners on Uniform State Laws recommended in 1993 that legislative history be demoted to a lower tier of sources. *See* UNIF. STATUTE & RULE CONSTR. ACT §§ 18–20, 14 U.L.A. 71–74 (1994); *id.* § 20 cmt. at 75 (confirming that the recommended "step by step" approach would reserve legislative history for cases where statutory meaning is otherwise "uncertain"). Like the Conference's 1965 effort at a uniform statutory construction act, which was not as clearly committed to lexical ordering, *see* UNIF. STATUTORY CONSTR. ACT §§ 13, 15 (withdrawn 1995), 14 U.L.A. 522–28 (1968) (offering a nonexhaustive list of sources that may be considered when a statute is ambiguous without expressly endorsing a plain meaning rule), few state legislatures followed the 1993 proposal.

60 TEX. GOV'T CODE ANN. § 311.023 (West 2017).

61 OR. REV. STAT. ANN. § 174.020(3) (West 2018) ("A court shall give the weight to the legislative history that the court considers to be appropriate."). Note the labor-saving device offered to judges, who were granted permission to limit review to sources "that the parties provide to the court." *Id.*

pretext of pursuing its spirit,"[62] and then approves consideration of legislative history and other listed sources "[w]hen the words of the statute are not explicit."[63] Several other state codes are equally or nearly as explicit about demoting legislative history.[64]

How much these statutes matter is open to question. Construction acts have not always influenced even official judicial policy,[65] and their effects on case results are unknown. Despite the openness to legislative history in the text of the Texas code, the Texas Supreme Court has declared that "legislative history cannot override a statute's plain words."[66] Last year, that court suggested that the matter is beyond the legislature's authority.[67] Nearly ten years ago, the Oregon Supreme Court saw its legislature's invitation to assign weight to legislative history and responded with a rule: "[N]o weight can be given to legislative history" when statutory text can have only one meaning.[68] That court might now be tacking back, away from lexical ordering.[69] Indeed, headstrong judicial leadership can run either way. New Mexico's code appears to reserve "the purpose of a statute . . . as determined from the legislative . . . history of the statute" for cases of residual uncertainty.[70] Yet New Mexico's Supreme Court has affirmed that "[t]he plain meaning rule 'must

62 1 PA. STAT. AND CONS. STAT. ANN. § 1921(b) (West 2018).

63 *Id.* § 1921(c)(7).

64 *See* LA. STAT. ANN. §§ 1:4, 24:177 (2018) (regarding letter and spirit, and legislative history for ambiguity); MINN. STAT. ANN. § 645.16(7) (West 2018) (same); N.D. CENT. CODE ANN. §§ 1-02-05, 1-02-39.3 (West 2018) (same); *see also* CONN. GEN. STAT. ANN. § 1-2z (West 2018) (referring to meaning that is "plain and unambiguous and does not yield absurd or unworkable results"); N.M. STAT. ANN. § 12-2A-20.C(2) (West 2018) (reserving "the purpose of a statute . . . as determined from the legislative or administrative history of the statute" for cases of residual uncertainty). The Colorado, Iowa, and Ohio construction acts have the second clause on legislative history without the first clause on letter and spirit. *See* COLO. REV. STAT. ANN. § 2-4-203(1)(c) (West 2018); IOWA CODE ANN. § 4.6.3 (West 2018); OHIO REV. CODE ANN. § 1.49(C) (West 2018).

65 *See* Gluck, *supra* note 12, at 1771–91 (examining state court opinions in five jurisdictions including Oregon and Texas, and showing apparent departures from statutory instructions for interpretation); Amy Widman, *Interpretive Independence: The Irrelevance of Judicial Selection and Retention Methods to State Statutory Interpretation*, 70 N.Y.U. ANN. SURV. AM. L. 377, 393–94, 401, 404–08 (2015) (counting such departures in a sizable minority of supreme court opinions in eleven states).

66 *In re* Collins, 286 S.W.3d 911, 918 (Tex. 2009).

67 *See* BankDirect Capital Fin., LLC v. Plasma Fab, LLC, 519 S.W.3d 76, 85 (Tex. 2017) ("Interpretive prescriptions, or permissions, to put a finger on the scale and stretch text beyond its permissible meaning invade the courts' singular duty to interpret the laws."); *see also* Billeaudeau v. Opelousas Gen. Hosp. Auth., 218 So. 3d 513, 516 n.4 (La. 2016) (similar for a construction act regarding demotion of legislative history).

68 State v. Gaines, 206 P.3d 1042, 1051 (Or. 2009).

69 *See* Spearman v. Progressive Classic Ins. Co., 396 P.3d 885, 888 (Or. 2017) (indicating that the court must "determine the meaning of the words of the statute most likely intended by the legislature that enacted it, taking into account its text in context and the relevant legislative history," yet citing *Gaines*); Lake Oswego Pres. Soc'y v. City of Lake Oswego, 379 P.3d 462, 468 (Or. 2016) (similar).

70 N.M. STAT. ANN. § 12-2A-20.C(2) (West 2018).

yield on occasion to an intention otherwise discerned in terms of equity, legislative history, or other sources.'"[71]

2. Agency Interpretations

A similar trend applies to agency positions. For federal courts dealing with federal agencies, part of the *Chevron* doctrine calls for lexical ordering with a clarity hinge. Federal courts are not supposed to defer to agency interpretations when the meaning of a federal statute is adequately clear on the issue at hand ("Step One").[72] Almost surely the doctrine means to stop the agency's position from influencing the judiciary's evaluation of statutory clarity; otherwise deference would creep into the decision of whether to defer.[73] There are other parts to the doctrine—only some agency interpretations are eligible for deference ("Step Zero"),[74] and apparently courts are supposed to apply some kind of reasonableness test when the statute is unclear ("Step Two").[75] Nonetheless, Step One imposes a clarity test that is analytically similar to the hinge for considering legislative history discussed above.

Chevron was decided in 1984, during the transition away from *American Trucking*'s inclusiveness. *American Trucking* itself cited not only legislative history but also an agency's position as factors in its decision.[76] Thus *Chevron* can be understood as displacing a multifactor test that was supposed to calibrate the degree of deference to the persuasiveness of the agency position, and then add other interpretive sources to the mix.[77] *Chevron* was unanimous among the seven Justices participating, and we might wonder whether part of the doctrine's attractiveness was the joinder of plain meaning with a second-tier consideration. Each type of source retained potential influence on outcomes, even though the analysis seemed to be structured differently. Either way, *Chevron* was a forerunner to a broader set of lexical ordering moves at the Court.

State-level endorsement of *Chevron*-style lexical ordering is common, but outliers do exist. In approximately five western states, judges are supposed to give no deference or weight to a home state agency's interpretation of a

71 Fowler v. Vista Care, 329 P.3d 630, 633–34 (N.M. 2014) (quoting Sims v. Sims, 930 P.2d 153, 157 (N.M. 1996)).

72 *See, e.g.*, Kingdomware Techs., Inc. v. United States, 136 S. Ct. 1969, 1979 (2016); Chevron, U.S.A., Inc. v. Nat. Res. Def. Council, 467 U.S. 837, 842 (1984).

73 *See* Philip Morris USA, Inc. v. Vilsack, 736 F.3d 284, 289 (4th Cir. 2013) (stating that, at *Chevron* Step One, "the court gives no weight to the agency's interpretation"). A different view is argued in Strauss, *supra* note 29, at 1164–65.

74 *See* United States v. Mead Corp., 533 U.S. 218, 227–31 (2001); Thomas W. Merrill & Kristin E. Hickman, Chevron*'s Domain*, 89 GEO. L.J. 833, 873 (2001).

75 *See Chevron*, 467 U.S. at 843–45, 851, 865–66; *cf.* Matthew C. Stephenson & Adrian Vermeule, Essay, Chevron *Has Only One Step*, 95 VA. L. REV. 597, 599–603 (2009) (contending that Step Two can be folded into Step One or otherwise made redundant with arbitrariness review).

76 *See* United States v. Am. Trucking Ass'ns, 310 U.S. 534, 545 (1940).

77 *See* Manning, *supra* note 30, at 162–63; Thomas W. Merrill, *Judicial Deference to Executive Precedent*, 101 YALE L.J. 969, 972–73 (1992).

home state statute.[78] Additionally, Iowa courts demand that a state statute clearly delegate discretion to the state agency on the disputed issue before entertaining deference to the agency's interpretation.[79] On the other hand, a few state courts indicate that an agency interpretation is one factor in the overall mix of considerations, or that the agency's position is considered in determining the plainness of statutory meaning.[80]

Nonetheless, most state doctrine accords with *Chevron* on the priority of clear statutory meaning. In recent years, approximately thirty-five state supreme courts have more or less plainly announced that they support an analytical step in which judges independently determine whether a statute is sufficiently clear to foreclose deference or weight to an agency interpretation.[81] In Massachusetts, for example, courts tell us that they review an agency's statutory interpretation de novo—but they also depict an analytical structure that largely tracks *Chevron*, asking first "whether the Legislature has

78 *See* Golden Rule Ins. Co. v. Tomlinson, 335 P.3d 1178, 1188 (Kan. 2014) ("No deference is paid to an agency's statutory interpretation."); Mont. Dep't of Revenue v. Priceline .com, Inc., 354 P.3d 631, 636 (Mont. 2015) (reviewing for correctness); Aline Bae Tanning, Inc. v. Neb. Dep't of Revenue, 880 N.W.2d 61, 65 (Neb. 2016) (asserting independent review); Midwest Railcar Repair, Inc. v. S.D. Dep't of Revenue, 872 N.W.2d 79, 85 (S.D. 2015) (rejecting deference for conclusions of law); Seherr-Thoss v. Teton Cty. Bd. of Cty. Comm'rs, 329 P.3d 936, 943 & n.2 (Wyo. 2014) (similar, though leaving room for deference to longstanding agency interpretations). There have been equivocations. *See* Sierra Club v. Mosier, 391 P.3d 667, 684–85 (Kan. 2017) (attempting to distinguish situations where the statute grants the agency a zone of discretion); Bostwick Props., Inc. v. Mont. Dep't of Nat. Res. & Conservation, 296 P.3d 1154, 1159 (Mont. 2013) (listing agency construction as one of four "factors" in statutory interpretation for cases involving the Department of Natural Resources); Project Extra Mile v. Neb. Liquor Control Comm'n, 810 N.W.2d 149, 162–63 (Neb. 2012) (acknowledging occasional court statements supporting agency weight, especially after the legislature fails to amend); Pub. Serv. Comm'n v. Qwest Corp., 299 P.3d 1176, 1182 (Wyo. 2013) (articulating a *Chevron*-like formulation).

79 *See* Myria Holdings Inc. v. Iowa Dep't of Revenue, 892 N.W.2d 343, 347 (Iowa 2017); *cf.* IOWA CODE ANN. § 17A.19.11.b–c (West 2018) (addressing the test for agency deference, without the word "clearly").

80 *See, e.g.*, N.M. STAT. ANN. § 12-2A-20.B(4) (West 2018) (listing administrative construction in the top tier); Marbob Energy Corp. v. N.M. Oil Conservation Comm'n, 206 P.3d 135, 139 (N.M. 2009) (stating that the court is merely "less likely to defer" when the statute is clear).

81 Authorities are on file with the author. Twenty-seven state supreme courts, in my view, plainly support this approach. In eight other states, the supreme court's current position is not as easy to code but there are noteworthy indications of support for such ordering. *See, e.g.*, Motor Vehicle Admin. v. Krafft, 158 A.3d 539, 547 (Md. 2017) (stating that courts "may accord some weight" to an agency's statutory interpretation without adverting to statutory clarity); Md. Ins. Comm'r v. Cent. Acceptance Corp., 33 A.3d 949, 958 (Md. 2011) (stating, six years earlier, that "when the language of the statute is clear and unambiguous, no deference is due"). Other helpful reviews of state court positions on agency deference include ESKRIDGE, *supra* note 16, at 283–88, and Bernard W. Bell, *The Model APA and the Scope of Judicial Review: Importing* Chevron *into State Administrative Law*, 20 WIDENER L.J. 801, 818–19 (2011).

spoken with certainty on the topic in question."[82] Like their treatment of legislative history, state supreme courts in the mid-1970s were less likely to explicitly demote agency interpretations. During those years, the range of state court positions was more balanced.[83]

* * *

Just because a court endorses lexical ordering hardly ensures that judges will do it. Some federal judges might think that certain elements of interpretive method are matters of personal choice,[84] and, even when they try, judges might not always follow instructions. Even high court opinions may backslide on the orthodox approach. *King v. Burwell*[85] references congressional committee work—albeit quickly and bashfully—to develop a background understanding of what the Affordable Care Act was trying to achieve and how.[86] Careful examination of the Act's text appears later on.[87] Regardless, neither officially prescribed nor actually practiced interpretive method is fully stable over time. Official support for lexically ordered statutory interpretation has risen and fallen before.[88]

Furthermore, even if implemented flawlessly, lexical ordering leaves open many choices. Judges still must decide which sources belong in which tiers, how to weight and interact sources within the same tier, and the content of the doctrinal hinge between tiers. The proper weighting and interaction of interpretive sources are old challenges that do not go away when lexical ordering is adopted. The assignment of sources to tiers and the construction of a hinge based on "clarity" are additional tasks that are unneeded for outright exclusion or inclusion of sources.[89]

---

82 Biogen IDEC MA, Inc. v. Treasurer & Receiver Gen., 908 N.E.2d 740, 750 (Mass. 2009) (quoting Goldberg v. Bd. of Health of Grandby, 830 N.E.2d 207, 213 (Mass. 2005)) (citing Chevron U.S.A., Inc. v. Nat. Res. Def. Council, 467 U.S. 837 (1984)); *see also* Peterborough Oil Co. v. Dep't of Envtl. Prot., 50 N.E.3d 827, 832 (Mass. 2016) (similar).

83 Using the same search queries for 1973–77, eleven state supreme courts more or less plainly endorsed lexical inferiority for agency interpretations, nineteen courts more or less plainly rejected it in favor of mixing considerations or de novo review, and seventeen courts did not offer an opinion on the issue. Authorities are on file with the author.

84 *Cf.* Evan J. Criddle & Glen Staszewski, Essay, *Against Methodological Stare Decisis*, 102 GEO. L.J. 1573, 1576–77 (2014) (asserting that federal courts tend not to give stare decisis effect to interpretive methodology for statutes); Gluck & Posner, *supra* note 16, at 1316, 1345–46 (indicating that the interviewed circuit judges accepted *Chevron* doctrine as binding, and that many interviewees considered legislative history if statutory text is unclear, but that several interviewees also doubted that the Supreme Court may impose broad interpretive rules beyond certain canons).

85 135 S. Ct. 2480 (2015).

86 *See id.* at 2486 (citing a Senate committee hearing in an introductory discussion).

87 *See id.* at 2489–90 (concluding that the Act was unclear on the issue of subsidy availability). The sequence of analysis in a written opinion is not necessarily the sequence for the decision process, of course.

88 *See supra* text accompanying notes 37–44.

89 For complaints that the clarity test lacks guiding content, see *supra* notes 15–17. Courts have disputed over which sources belong in which tiers. *See, e.g.*, Yates v. United States, 135 S. Ct. 1074, 1081–83 (2015) (plurality opinion) (relying on a statutory title); *id.*

Indeed we should entertain the possibility that lexical ordering is a low-stakes choice. Its logical structure might be sufficiently unstable, ignored, narrow, manipulable, or flexible across judges with different interpretive inclinations to make little difference in adjudication. But these are possibilities rather than established facts. Consider that judges act as if lexical ordering is a distinctive and meaningful option—endorsing it in many, but not all, jurisdictions and for many, but not all, fields of legal interpretation. So prudence also recommends that we take seriously the widespread official position on lexical ordering for statutory cases, and think hard about whether some form of it can be justified and implemented.

II. Trade-Offs in Theory

Trends in statutory interpretation are subject to change, as earlier proponents of lexical ordering found out. If and when a different trend emerges, judges probably will act as central players in reform efforts, as they have in the past. In any event, a trend is not a justification. And today we should wonder whether lexical ordering's justifications match its current popularity. This Part identifies core advantages and disadvantages as a matter of general theory. The discussion begins with a kind of ideal decision situation that assumes away decision costs, indecision risks, and any need for judges to compromise, then foregrounds such considerations to give lexical ordering a better defense. The goal here is to advertise major trade-offs without forcing definitive conclusions. With little existing work on the deep questions for lexical ordering in statutory cases, we have to take initial steps that leave some matters unresolved.[90]

That said, much of the analysis generalizes across sources and goals for statutory interpretation. In fact, similar trade-offs apply whenever judges try to categorically prioritize bundles of considerations that are part of the same decision.[91] So, too, for structures well beyond adjudication, such as the

---

at 1094 (Kagan, J., dissenting) (asserting that statutory titles cannot limit plain meaning in operative provisions); Philip Morris USA, Inc. v. Vilsack, 736 F.3d 284, 289 (4th Cir. 2013) (indicating that legislative history should be considered at *Chevron* Step One); United States v. Geiser, 527 F.3d 288, 294 (3d Cir. 2008) (indicating that legislative history should be considered at Step Two). Arguably, judicial precedent sits in the highest tier of sources. *See* Kimble v. Marvel Entm't, LLC, 135 S. Ct. 2401, 2409 (2015) (reiterating that stare decisis is supposed to have "enhanced force" for cases interpreting statutes).

90 The analysis below concentrates on *source categories* (e.g., legislative history, agency interpretations), rather than particular items that are examples of such sources (e.g., the committee report in *Train*, the agency position in *Chevron*). Today's judicial instructions for lexically ordered statutory interpretation rest on such broad categories. *See supra* Section I.B. Also note that lexical ordering sometimes has more than two tiers. The analysis below concentrates on a two-tier structure, which is the least complicated version and a sensible starting point.

91 *See* Samaha, *Tiebreakers*, *supra* note 13, at 1673 (investigating "lexically ordered rules that address the same decision," while pointing out that "there is no orthodox way to identify the relevant 'decision'"). For a challenge, think about whether subject-matter jurisdic-

proper relationships among law, ethics, and private ordering,[92] and the evaluation of political systems.[93] In any of these fields we are free to ask about the justifications for demoting considerations, the effect on decision quality, the likelihood that decisionmakers will ignore lower-tier sources, and so on. Of course, each application takes place in its own decision environment. We take up statutory interpretation here. If this analysis is cast too narrowly, there will be opportunities for expansion later on.

A. Ideal Decision Situations

Suppose, unrealistically, that interpretation is costless and that interpreters face no problem resolving cases decisively or agreeing on a general method of decision. In such ideal decision situations, lexical ordering becomes not only odd but troubling.

1. Against Lexical Ordering

The core risk is valuable information loss. If a lower-tier source is relevant to the decision, lexical ordering entails that decisionmakers lose whatever value it has for the fraction of decisions in which top-tier sources are deemed sufficient. To illustrate and simplify, if judges conclude that top-tier sources are sufficiently clear about 50% of the time, but find lower-tier sources useful about 50% of the time, then about 25% of all decisions will suffer the exclusion of useful information.[94] In this quarter of the docket, lower-tier sources have no power to make the call closer or easier than it first seemed, or to influence the formulation of governing doctrine, or to flip the final judgment. We may characterize any of these effects as losses in decision quality, especially the latter two. And the threat holds regardless of one's values or goals for statutory interpretation.

It is true that lower-tier information presumably is less valuable than top-tier information.[95] Perhaps decisionmakers can select lower-tier source cate-

---

tion and the so-called merits of a case should count as one lexically ordered decision about case outcome, or two different decisions that merely share the same decision tree.

92 *See id.* at 1717–37 (considering law as life's tiebreaker, and vice versa).

93 *See* RAWLS, *supra* note 26, at 42–43. Under the Rawls framework, people deserve a base level of various goods before anyone is made substantially better off than others. *See id.* at 43, 541–42 (referring to a certain base level of wealth alongside basic liberties). These priorities were grounded on what Rawls thought would be accepted in his hypothetical original position decision situation and upon our considered judgment, but the lexical ordering in Rawls's theory was not extensively defended. *See id.* at 542–48. It might be imprudent to implicate his work on this score, especially as an aside. But we should be sensitive to what lexical ordering demands and be careful to put that logical structure in its proper place.

94 0.5 * 0.5 = 0.25, assuming that the probability of lower-tier sources being useful is independent of the top-tier sources seeming adequate. These numbers are loosely based on results from the survey described in Section III.B.

95 *See* Samaha, *Tiebreakers*, *supra* note 13, at 1699–700 (suggesting that the most valuable sources be placed in the top tier, all else equal); *cf.* Louis Kaplow, *Optimal Multistage*

gories such that the probability of their usefulness is higher than zero yet far lower than 50%. (If a source category has zero chance of being useful, the category should be flatly excluded, not lexically ordered.) As well, lowering the probability of stopping with the top-tier sources, such as by loosening up what counts as "unclear," will decrease the risk of information loss. Decisionmakers also may load up the top tier with as many potentially useful sources as possible. But any of the foregoing options will merely lower the probability that lower-tier source information will be useful or will be ignored when it is. Just yet, we have no reason to tolerate *any* such risk. When a lower-tier source turns out to make no difference, no harm is done by considering it—not on our initial assumption of costless decisionmaking. The quality losses from lexical ordering are deadweight.

These simple thoughts are consistent with more sophisticated normative decision theory. One straightforward position on information in ideal decision situations is that "more information is better."[96] Assuming relevance and barring cognitive difficulties, increasing the amount of information considered should increase the chance of an accurate, reliable, and otherwise high-quality decision. More formally, the principle of total evidence for rational actors recommends the use of all available evidence when estimating the likelihood of various outcomes—again, bracketing the costs of collecting and using the information.[97] One can make the same claim about any source that is relevant to understanding and applying statutes in adjudication.

Inclusive theoretical approaches to information leave no apparent room for lexical ordering, but they do allow us to assign different weights and confidence levels for different considerations. By way of analogy, rational choice and expected utility maximization theorists long ago developed weighted-additive models for integrating a large number of decision factors.[98] Weights can follow the estimated reliability of broad source categories or narrower subsets thereof. Ordinary meaning of statutory text could be assigned very heavy influence, statutory titles could get little but not zero weight, while

Adjudication, 33 J.L. ECON. & ORG. 613, 644–45 (2017) (suggesting that adjudication stages might be sequenced such that information with "a high ratio of diagnosticity to cost" is gathered first).

96 *See* J. Edward Russo, *More Information Is Better: A Reevaluation of Jacoby, Speller and Kohn*, 1 J. CONSUMER RES. 68, 71 (1974) (questioning an earlier study of information overload).

97 *See* I.J. Good, *On the Principle of Total Evidence*, 17 BRIT. J. FOR PHIL. SCI. 319, 319 (1967).

98 *See, e.g.*, RALPH L. KEENEY & HOWARD RAIFFA, DECISIONS WITH MULTIPLE OBJECTIVES 137–38 (1993); JOHN W. PAYNE ET AL., THE ADAPTIVE DECISION MAKER 24 (1993). A similar analogy can be made to Bayesian updating. Beliefs about the probability of a proposition being true can be adjusted based on a stream of new information, with different items having different effects. *See generally* Michael O. Finkelstein & William B. Fairley, *A Bayesian Approach to Identification Evidence*, 83 HARV. L. REV. 489, 497–500 (1970) ("We tend to see a case as a whole . . . .").

committee reports and agency positions could fall somewhere in between.[99] On similar logic, we have to distinguish lexical ordering from mere presumptions. Presumptions remain within the overall mix of considerations, whatever their weight.[100] A lexical ordering of considerations cannot give that assurance of inclusive evaluation.

At the same time, thoughtful decisionmakers might flatly exclude troublesome considerations. In life and law, people try to avoid relying at all on disfavored bases for particular decisions, even when the information is cheap to process. For instance, many people want to ensure that in most circumstances job applicants are not hired because of their race, religion, or sex. A combination of moral, ethical, social, and legal commitments steer people away from reliance on particular reasons for particular decisions. Lexical ordering cannot satisfy such commitments to exclusion either—including the opposition to legislative history as a bad substitute for statutory text, and the opposition to deference toward agencies as an illegitimate shift in power away from judges. Lexical ordering is rule-like in excluding information only after judges decide that other sources are clear enough.

Nor is lexical ordering a conventional evidentiary form of conditional relevance. Information sometimes is deemed irrelevant in the sense of having no value in helping to demonstrate a proposition of interest because some other proposition is not possibly or not likely enough true. Conditional relevance is part of the law of evidence, which suggests that "when a spoken statement is relied upon to prove notice to X, it is without probative value unless X heard it."[101] This notion has companions in ordinary decisionmaking. Thus a rational employer may conclude that job applications should not be evaluated unless the employer has adequate revenue to hire. The basic idea is that certain information is not at all helpful unless a specified condition is met or assumed.

---

99 On the usefulness of plain meaning as a coordination device, without necessarily defending lexical ordering, see David A. Strauss, Essay, *Why Plain Meaning?*, 72 NOTRE DAME L. REV. 1565, 1565 (1997), and Frederick Schauer, *Statutory Construction and the Coordinating Function of Plain Meaning*, 1990 SUP. CT. REV. 231, 232, 250–56. Compare also the conventional rankings of influence for various subcategories of legislative history. *See* KATZMANN, *supra* note 31, at 54.

100 *See* Sebelius v. Auburn Reg'l Med. Ctr., 568 U.S. 145, 156 (2013) (stating that canons of construction are merely rules of thumb that "can tip the scales when a statute could be read in multiple ways"). I have set aside so-called conclusive presumptions that function as rules of exclusion.

101 FED. R. EVID. 104 advisory committee's note to the 1975 proposed rules; *see* FED. R. EVID. 104(b) (requiring some proof of a fact if the relevance of evidence depends on that fact). Rule 104(b) does not spell out the dimensions of the idea, however, which seems difficult to separate and reconcile with the rules' broad notion of "relevance." *See* Ronald J. Allen, Essay, *The Myth of Conditional Relevancy*, 25 LOY. L.A. L. REV. 871, 877 (1992) ("No evidence is simply relevant in its own right."); Richard D. Friedman, *Conditional Probative Value: Neoclassicism Without Myth*, 93 MICH. L. REV. 439, 441–45 (1994) (illustrating classical conditional relevance problems and the argument that such evidence is relevant under the rules without evidence of the predicate, if the predicate is at least possibly true).

The evidentiary idea is not enough to sustain the reach of lexical ordering in statutory cases today, however, as William Baude and Ryan Doerfler point out.[102] Nobody seems to maintain that legislative history and agency interpretations cannot be "relevant," in an evidentiary sense, to the issue of how to understand or apply statutes *unless* the statutory text is otherwise unclear. Judges certainly can evaluate whether a committee report or an agency's view actually addresses the disputed issues in a case regardless of whether other sources make the statute seem clear or unclear. Theoretically, judges need not make any estimate of statutory textual clarity to decide whether such lower-tier sources have at least some value in understanding or applying a given statute.[103] It is just that many people contend that otherwise clear statutory text should stop the analysis *whether or not* lower tier sources speak directly to and support a different statutory meaning and a different judgment. That contention draws on a sense of propriety beyond evidentiary relevance, and it should be grounded in a thoughtful normative theory of interpretation.

2. Principled Beginnings

The theory is hard to build. It cannot rest on the conclusion that, as some courts have put it, "the best" evidence of legislative intent or statutory meaning resides in a particular set of sources.[104] However true, that conclusion just does not recommend the multitiered structure at issue. Identifying

102 *See* Baude & Doerfler, *supra* note 13, at 540, 547–49 (arguing that, at least ordinarily, evidence of statutory meaning is either relevant or not, rather than relevant only if meaning is otherwise unclear). Below, I take up the related yet distinct idea that higher-tier and lower-tier sources implicate different judicial activities. *See infra* text accompanying note 113. In pursuing that line of argument, one might say that lower-tier sources are not "relevant" to what the judge should do with the top-tier sources, because the two separate categories of sources match up with two separate tasks. But conditional relevance issues in evidence do not seem to follow that logic.

103 The point is easier to make with legislative history, and the conclusion depends on the rationale for considering the lower-tier source. Consider, for example, the argument that legislative history is worthless to the extent that legislators are unaware of it. This argument does depend on a factual predicate (unawareness) but not on statutory clarity. Compare the argument that an agency interpretation is not valuable to the extent that the agency has not been or may not be delegated interpretive authority by the legislature. This argument depends on a predicate (delegation) that *can* be logically connected to statutory clarity, if we adopt a particular view about the foundations of deference: Some sources suggest that deference is based on signs of legislative intent, *see, e.g.*, United States v. Mead Corp., 533 U.S. 218, 229–31 (2001) (referring to congressional expectations), and thus statutory clarity could be taken to indicate that the legislature did not intend to delegate interpretive authority to the agency. But an alternative understanding is that legislative intent is nonexistent on the matter of delegation and that deference is justified on institutional grounds. *See, e.g.*, Eric A. Posner & Cass R. Sunstein, Chevron*izing Foreign Relations Law*, 116 YALE L.J. 1170, 1193–94 (2007).

104 *E.g.*, People v. Bradford, 50 N.E.3d 1112, 1115 (Ill. 2016); *see* Scher v. Burke, 395 P.3d 680, 685 (Cal. 2017) (indicating that statutory text is generally the most reliable source for legislative intent).

the best source is no more supportive of lexical ordering than of assigning heavy weight to that source within a mix of considerations.[105] After all, *American Trucking* singled out statutory text as unsurpassed in importance for ascertaining legislative purpose.[106] Lexical ordering didn't follow.

The principled proponent's task is unenviable indeed. It is to show that top-tier sources are so important that lower-tier sources should have zero effect on the decision no matter how persuasive the latter might be after thoughtful consideration, unless the former are, in isolation, judged inadequate for decision—and yet the lower-tier sources must not be so unimportant or problematic that they should be excluded flatly. This needle might be threaded with sources that are morally fraught to the point of being nearly illegitimate, which could be shunted aside for a last resort when other sources are truly exhausted.[107] Without extra motivation, however, such as the need to economize on decision costs or clear dockets, this kind of principled justification must be rare.

For statutory interpretation, John Manning began the task of justifying the demotion of purpose informed by legislative history. He did not argue decision costs or indecisiveness.[108] Instead Manning contended that one plausible version of purposivism itself would rule out the old notion of "the spirit" of the law trumping law's text, in favor of the position that clear statutory text trumps evidence of background purposes.[109] This version of purposivism seeks to respect legislative choices between specific rules that tightly constrain the discretion of interpreters and vague standards that effectively delegate some implementation authority, while "enabl[ing]" Congress to make such choices via statutory text.[110] Manning pointed to endorsements of clear statutory text's primacy by judges who do not endorse hardline textualism, and he admirably explored lexical ordering as a matter of normative legal theory beyond pragmatic compromise.

Manning could be right about purposivism, and his work is thoughtful, but the case for lexically ordered interpretation is incomplete. Much of Manning's analysis is effective against purpose trumping statutory text as understood with a limited set of sources; and this trumping might well hinder

105 *See* Kimmel v. State, 80 N.E.3d 370, 377 (N.Y. 2017) (observing that "the plain language of the statute provides the best evidence of legislative intent" but without making legislative history lexically inferior).

106 *See* United States v. Am. Trucking Ass'ns, 310 U.S. 534, 543 (1940) ("There is, of course, no more persuasive evidence of the purpose of a statute . . . .").

107 *See* Samaha, *Tiebreakers*, *supra* note 13, at 1677–80.

108 *See* Manning, *supra* note 30, at 147–65 (stressing institutional settlement and constitutional law).

109 *See id.* at 115–16, 129–32.

110 *See id.* at 147–48, 163–65; *see also* John F. Manning, *What Divides Textualists from Purposivists?*, 106 COLUM. L. REV. 70, 96 (2006) ("[T]extualists believe that judicial adherence to semantic detail (when clear) is essential if one wishes legislators to be able to strike reliable bargains.").

Congress from using statutory text to make implemental choices.[111] But the *Holy Trinity*–style position under attack is a version of lexical ordering, and knocking out the lexical superiority of purpose does not support the lexical superiority of any other source. The options for judges are not only which sources to lexically order, but whether to lexically order at all. Worth heavy emphasis is that sources thought to be highly valuable (such as ordinary meaning or judicial precedent) can be assigned extremely heavy weights in an all-things-considered judgment, while sources considered less valuable (such as floor debates or arcane canons) can be assigned extremely low weights. Or excluded.[112] Lexically ordered interpretation should perform better than these alternatives.

Another line of justification begins by associating different sources with fundamentally different judicial activities. For instance, we could acknowledge that judges must decide statutory cases somehow, yet distinguish judgments that rest on a narrow conception of statutory interpretation over judgments that rest on additional grounds.[113] Sources in every tier could be useful in closing cases, but perhaps only sources in the top tier would be thought useful in understanding the true meaning or intent of statutory text. Lenity or a canon favoring veterans might be persuasively characterized as useful for closing cases but not for understanding the meaning or intent of enacted statutory language,[114] while legislative history might be more difficult to separate on this basis. Wherever the lines fall, the idea of partitioning different kinds of sources into different categories of use is entirely logical.

All of that is a hopeful beginning but, again, stops short of a convincing justification. Standing alone, fundamental differences do not recommend lexical ordering. People regularly face the complication of relevant consider-

111 Judges also must determine *how* Congress has chosen to convey its implemental choices, whether through ordinary meaning of enacted text, legislative history, or a combination. Placing any source in a lexically superior tier might unjustifiably disable Congress's ability to convey messages in more than one way (and to divide its labor). For discussion of ex ante effects on the legislative drafting process, see notes 128 and 143 and accompanying text.

112 A very low weight for a source category will make that category relatively uninfluential to decisions and thus possibly not worth the effort. But that idea relies on decision costs.

113 My thanks to Caleb Nelson and Ryan Doerfler for help in formulating this point. For a relatively narrow conception of interpretation that is compatible with the notion in the text above, see Lawrence B. Solum, Essay, *The Interpretation-Construction Distinction*, 27 CONST. COMMENT. 95, 104 (2010). Solum connects interpretation to evidence of language usage. *But cf. id.* at 103 n.19 (slotting judicial application of legal norms into the category of construction). For a broader conception of interpretation, see Kent Greenawalt, *Constitutional and Statutory Interpretation*, *in* THE OXFORD HANDBOOK OF JURISPRUDENCE AND PHILOSOPHY OF LAW 268, 268–70 (Jules Coleman & Scott Shapiro eds., 2002). Greenawalt uses the term to include, among other things, application to particular cases and stare decisis.

114 *See* Anita S. Krishnakumar, *Reconsidering Substantive Canons*, 84 U. CHI. L. REV. 825, 833–34, 838–39, 878 (2017) (describing the conventional distinction between substantive and semantic canons, while observing that substantive canons and evidence of legislative intent might point in the same direction).

ations that seem different in kind or measured on different scales of value,[115] but lexical ordering is hardly the only response.[116] Even if different source categories initially must be thought about separately, decisionmakers may strive to reconcile or combine their implications before reaching a final decision. It happens every day, at least implicitly. Additional argument is needed to show that sources truly aimed at meaning, which sometimes will close cases, should not always be mixed together with sources truly aimed at closing cases. Both types of source would share an important function in closing cases, after all, regardless of whether their overall value is difficult to estimate and aggregate. In starkest form, lexical ordering requires us to accept that no amount of value in a lower-tier source will have any influence on decisions in which top-tier sources tilt perceptibly in one direction or another. Wanting a less rigid gateway to the lower tier suggests that those sources are not so very categorically inferior in the first place.

The other stumbling block for this justification is artificial, but it emphasizes the artificiality of ideal decision situations. Even if we assume a convincing basis for treating some sources as both different and fundamentally less important—which has not been ruled out—the present argument for lexical ordering depends on a problem of indecision. The suggestion is that lower-tier sources are categorically inferior grounds for judgment, yet they are reserved for situations in which top-tier sources, regrettably, cancel out. The lower tier is created to solve this problem. To distinguish relatively pure principle from practical need, however, we assumed away this risk. Judges in an ideal world need not reserve inferior sources for a last resort that will never be reached. So, to prop up lexical ordering in statutory cases, we should make our view of judicial decisionmaking less ideal, more realistic, and more hard-bitten.

B. Nonideal Constraints

Additional hope for lexical ordering arrives when we add the problems of decision costs and indecision risks. The former is a standard consideration in information economics and institutional design, while the latter problem is less widely flagged. Both provide rational reasons for decisionmakers to ignore relevant information at times, which is a price that lexical ordering exacts. The challenge is identifying domains in which lexical ordering efforts make convincingly good sense. The analysis below suggests that the problem of decision costs might support the lexical inferiority of legislative history, while the problem of indecision might do so for agency interpretations. The more realistic we get, however, the more provisional our support should be.

115 *See* Cass R. Sunstein, *Incommensurability and Valuation in Law*, 92 MICH. L. REV. 779, 796–98 (1994).

116 *See* Jeremy Waldron, *Fake Incommensurability: A Response to Professor Schauer*, 45 HASTINGS L.J. 813, 815–16 (1994) (distinguishing dilemmas posed by strong incommensurability problems from lexical priorities); *see also* Samaha, *Tiebreakers*, *supra* note 13, at 1677.

1. Decision Costs

Sensible decisionmaking accounts for not only decision quality but also the costs of reaching decisions at various levels of quality.[117] Although saying so might be uncomfortable for judges, who must offer final judgments in cases that affect the well-being of countless other people, at some point increases in judicial decision costs will not be worth the predicted gains. Likewise, we should care about upstream costs. Knowing that judges will or might consider a source category such as legislative history incentivizes lawyers to research and argue the source, and, perhaps, encourages interested parties to produce more of it in the first place.[118]

These old ideas generate fine reasons for judges to depart from an all-inclusive approach to information, but alone they cannot justify lexical ordering. First of all, some of today's lexically inferior sources are cheap to process. Agency interpretations, if not subject to elaborate reasonableness testing, give straightforward answers to litigated questions by definition. The rule of lenity is even cheaper.[119] Second, outright exclusion is a handy solution for serious decision costs, as Adrian Vermeule emphasizes.[120] If a source category that is associated with substantial average decision costs has a 1% chance of being useful to judges, then the expected value might well be low enough to justify a rule of exclusion.[121] If instead the chance of usefulness is closer to 50%,[122] the risk of valuable information loss might well be too high to ever ignore the source.

Space for lexical ordering can open up, however, if its doctrinal hinge is diagnostic. The probability of a source category's usefulness might vary across case type, and the test for using lower-tier sources might help tell the

117 *See, e.g.*, Avery Wiener Katz, Essay, *The Economics of Form and Substance in Contract Interpretation*, 104 COLUM. L. REV. 496, 524 (2004) (observing that a "broader context can be purchased—but only at a cost of time and trouble, and of exacerbating certain incentive problems"); Adam M. Samaha, *Looking Over a Crowd—Do More Interpretive Sources Mean More Discretion?*, 92 N.Y.U. L. REV. 554, 561–62 (2017) [hereinafter Samaha, *Looking Over a Crowd*] (invoking decision cost and decision quality trade-offs); Strauss, *supra* note 99, at 1565–66 (discussing the upside of agreement in low-stakes situations).

118 *See* ADRIAN VERMEULE, JUDGING UNDER UNCERTAINTY 110–12 (2006); Antonin Scalia & John F. Manning, *A Dialogue on Statutory and Constitutional Interpretation*, 80 GEO. WASH. L. REV. 1610, 1612 (2012) ("It's there because we use it.").

119 These sources must be assigned weight if included with a larger mix, which might be difficult. However, weighting might be done once categorically. Anyway, weights are implicitly assigned to source categories when they are allocated to lower tiers and a doctrinal hinge is built.

120 *See* VERMEULE, *supra* note 118, at 192–97. Much of what I write here is consistent with Vermeule's analysis. His objection to intermediate solutions, such as plain meaning rules, was based on the asserted instability and inefficacy of the doctrines. *See id.* at 195–96. Inefficacy of lexical ordering is an empirical question on which I report some evidence in Part III.

121 For arguments supporting exclusion of legislative history when the benefits are speculative, see VERMEULE, *supra* note 118, at 192–97.

122 *See supra* note 94; *infra* Section III.B (reporting results from a survey of appellate judges).

difference. Suppose that a source category is useful in about 25% of all cases, but this figure falls to 1% in one part of the docket and rises to 50% in the remainder. Sadly for the cause of quality, lexical ordering is rule-like in instructing judges to ignore the content of the lower-tier source at issue while deciding whether to exclude it. But happily the chance of the lower-tier source being useful partly depends on the persuasiveness of the top-tier sources alone.[123] When the top-tier sources are highly reliable and in accord, even a crystal clear sign from a source such as legislative history is very unlikely to influence the decision. When instead the top-tier sources are foggy and discordant, even a weak inference from a lower-tier source is likely to make a difference.

A good hinge will save decision costs when the chance of usefulness is intolerably low and will accept those costs when the chance is adequately high. The proper formulation for such a diagnostic test is not at all evident. But vaguely asking judges to determine whether a statute seems otherwise "clear" is not a poor start. Judges could operationalize clarity to mean "very unlikely, as a predictive matter and in general, that the lower-tier source category will make a meaningful difference to the decision." If judges can make this determination reliably and cheaply, and if considering the lower-tier source would add substantial decision costs, then lexical ordering might be justified. The better judges are at distinguishing the 1% from the 50% useful sources, the more likely that lexical ordering will outperform a quality-sacrificing rule of exclusion or a cost-maximizing rule of inclusion.[124]

Still, this leaves many key values unknown or unconfirmed. We have not chosen values for each source category or for judicial decisions that use more rather than less information. Such generalizations are unavoidably imperfect, bound to vary across the docket, and contestable based on disagreement over the appropriate judicial mission in statutory cases. Without a sense of the beneficial value of sources within various case types, however, judges cannot be sure whether including, excluding, or demoting a given source is sensible. Nor do we have an evidence-backed estimate of how well judges can

123 This is most obviously true when "useful" is restricted to final judgments, as opposed to estimating the difficulty of the decision or formulating doctrine.

124 From this perspective, lexical ordering resembles a "take the best, ignore the rest" heuristic in which people facing heavy information loads screen out options using one or two considerations. *See* Gerd Gigerenzer & Daniel G. Goldstein, *Reasoning the Fast and Frugal Way: Models of Bounded Rationality*, 103 PSYCHOL. REV. 650, 653 (1996); *cf.* Richard H. Thaler et al., *Choice Architecture*, *in* THE BEHAVIORAL FOUNDATIONS OF PUBLIC POLICY 428, 435–36 (Eldar Shafir ed., 2013) (describing elimination by aspects). For an example and a graphic rendering of a take-the-best algorithm that is consistent with lexical ordering, see SANJIT DHAMI, THE FOUNDATIONS OF BEHAVIORAL ECONOMIC ANALYSIS 1420–23 (2016). Screening may rule out the option that would have been found best after thorough evaluation. The heuristic might or might not be rational depending on decision costs, stakes, and other factors. *See id.* at 1427–28 (noting questions about the frequency with which people use particular heuristics, and the appropriate benchmarks for evaluating them); *cf.* Baude & Doerfler, *supra* note 13, at 551 (attempting to distinguish plain meaning rules from heuristics in which some options are ruled out without making a final decision).

identify—sight unseen—sources that are low value using a clarity test. The task will be difficult for source categories that are relatively unfamiliar, complex, or diverse in quality across cases. If judges operationalize the clarity test loosely, to ordinarily allow consideration of lower-tier sources that might be useful, then less cost will be saved. There are no easy answers.

The decision-cost side of the analysis might be clearer, though. Little direct cost can be avoided by ignoring a single simple canon like lenity (or severity), and the cost implications of deference to agencies depend on how carefully judges analyze those positions. If judges routinely defer to agencies without serious reasonableness testing, then agency interpretations are not promising candidates for lexical inferiority on a decision-cost basis.[125] In contrast, ignoring legislative history surely avoids substantial decision costs in a fraction of cases. Occasionally the arguably relevant legislative history is new to the lawyers and judges, tens of thousands of words long, and requires hours of work to understand, integrate, and contextualize thoughtfully.[126] If one assigns legislative history a low but nonzero value, then it becomes a plausible candidate for lexical inferiority, assuming that the doctrinal hinge is sufficiently diagnostic.

On the other hand, the work need not be done from scratch in each case. Research detailed in a court opinion should help adjudicate or settle future cases. Plus legislative history is a large category. Subcategories of high-value and low-cost history might be identified for the top tier—perhaps past enactments and committee reports[127]—with the rest flatly excluded. And we should seriously doubt that lexical ordering has a unique ex ante effect on the legislative process or litigation. True, a court's firm commitment to the lexical inferiority of a source category should affect the probability of its influence on case results, as estimated by legislators, lawyers, and other interested observers. But theoretical effects on probability estimates are not special to lexical ordering commitments; reweighting sources

125 A quick Step Two check for a reasonable interpretive choice by the agency, in accord with conventional interpretive arguments, will cost judges little. Judges will have done the heavy analytical lifting at Step One. A simple rule of deference to agencies for unclear statutes will cost judges approximately nothing extra to execute. We should add any uncounted social costs associated with agencies producing these interpretations in the hope of getting deference, but often the agency will produce a litigating position regardless.

126 *See* VERMEULE, *supra* note 118, at 110 (highlighting examples). The analysis above simplifies matters by not breaking out the risk of judges erring in their understanding of legislative history.

127 *Cf.* Schwegmann Bros. v. Calvert Distillers Corp., 341 U.S. 384, 395 (1951) (Jackson, J., concurring) (contending that judges should only look to committee reports when resorting to legislative history); Jorge L. Carro & Andrew R. Brann, *The U.S. Supreme Court and the Use of Legislative Histories: A Statistical Analysis*, 22 JURIMETRICS J. 294, 304 tbl.2 (1982) (indicating that, from 1938 to 1979, about 45% of the Supreme Court's legislative history citations were committee report references). Carro and Brann's table shows over 3300 citations to House and Senate committee reports out of 7465 total legislative history citations.

can have the same effect.[128] Either way, we lack good measures of the lawyer, staff, and judge time needed to process legislative history across case types, let alone a settled number for what a judge's time is worth. Without more, lexical ordering remains a debatable response to the decision costs associated with only some of the sources that are demoted by judges today.

2. Indecision Risks

There is another reason for lexical ordering, however: avoiding intolerable ties.[129] Sometimes decisionmakers find that two or more options are equal on a given metric, or equal as far as they can tell with any confidence, and yet one option must be selected. For their part, judges are asked to reach decisive resolutions, without fail, even for the most difficult issues. During their careers, they will face some number of awfully close calls that are practically no different from perfect ties. And, as it happens, reserving a source in a lexically inferior position tends to drive down the chance of all sources canceling out, all else equal. This finding rests on several simplifying but not entirely unrealistic assumptions.

To illustrate, suppose that each source may strongly or weakly support either side's position in a case, or be unclear. Suppose further that each source is equally likely to bear any one of these five implications, and that the judge will add up the implications of all sources considered.[130] If all sources cancel out—the unclear sources are no help, and the strong and weak implications for one side offset the strong and weak implications for the other side—then the judge remains saddled with uncertainty. Adjudication stalls

128 Judges flatly excluding legislative history could have an effect on the incentives for interested parties to produce it, shape it, collect it, and substitute it for enacted text. *Cf.* Vermeule, *supra* note 17, at 94–95 (noting such theories and scarce evidence). One might then guess that judges taking an all-inclusive approach would have the opposite incentive effects, while the effects of lexical inferiority would lie somewhere in between.

But from an ex ante perspective, *both* lexical inferiority and a rule of inclusion generate positive probabilities of legislative history influencing court decisions, and the probabilities do not really depend on which approach judges endorse. An all-inclusive approach requires source weights, explicit or implicit; less weight can be assigned to disfavored source categories, which effectively reduces the probability that such sources will influence court decisions. Lexical ordering likewise affects the probability of influence among source categories, albeit using distinctive doctrinal tools: depending on the doctrinal hinge, among other factors, lexical ordering generates a probability that judges will consider and follow lower-tier sources—which is like lowering their expected weight. With lexical ordering, drafters theoretically can drive down the chance of judges considering legislative history to zero by writing statutes very clearly. But drafters should be able to approximate that result with an all-inclusive approach, if legislative history is lightly weighted or if drafters may instruct judges to ignore legislative history. Drafters might or might not believe that judges will follow the officially endorsed approach, and that the official approach is sufficiently specific to guide judges, but these are issues for all of the competing approaches on which more evidence is needed.

129 *See* Samaha, *Tiebreakers*, *supra* note 13, at 1685–1700.

130 We might hope that the relevant sources will not tend toward equipoise. But we are modeling litigation, which often involves represented parties contesting legally hard cases.

or turns lawless without more. One response is to increase the number of sources considered within the system. Whatever conventional wisdom might suggest, adding sources will increase decision costs but also should increase decision quality and, more to the point, will tend to *increase* decisiveness in this model.[131] On the above assumptions, moving from one to five such sources reduces the probability of an overall tie from 20% to about 12%.[132]

An alternative is lexical ordering. Judges may demote one source into a lexically inferior tiebreaking position, to be considered only when the other source inferences add up to a tie. Doing so presumably will yield somewhat lower quality decisions but can save decision costs and, more to the point, will tend to drive down the chance of a tie even *faster* than mixing all sources together. On the above assumptions, moving from one to five sources with one source reserved for tiebreaking reduces the probability of an overall tie from 20% to under 3%.[133] If judges choose a tiebreaker that always supports one side or another without ever being unclear itself, then the risk of an overall tie falls to zero.[134]

Thus, for promoting decisiveness, agency interpretation is a better candidate for the lower tier than legislative history—the opposite implication from our analysis of decision costs. Litigated agency interpretations are more decisive than legislative history, as a category, in the sense of plainly favoring a unique case result, as long as judges do not complicate matters with serious reasonableness testing. Legislative history often requires significant thought without a guaranteed implication in one direction or another. The rule of lenity is an even cleaner tiebreaker, offering unmatched speed and decisiveness once considered. Lenity can close every hard criminal case, agency interpretations can close hard civil cases when a deference-eligible agency position is available, and legislative history holds some chance of closing the remainder of the civil docket.[135]

---

131 *See* Samaha, *Looking Over a Crowd*, *supra* note 117, at 572–75 (relying on permutation tables and simulations). The allied intuition is that tossing lots of sources together, like tossing lots of coins, is not likely to yield an exactly equal number of implications pointing in different directions. *See id.* at 575. On the phenomena of judges cherry picking and spinning sources, see *id.* at 582–89.

132 *See id.* at 572, 573 fig.1, 574 fig.2. Decisionmakers cannot always add information, sensibly or at all. First, decision costs tend to rise. Second, decision quality might actually suffer at some point if the decision protocol becomes complicated or loaded with information. *See id.* at 590–93. Third, the number of relevant sources has an upper bound. People can develop new understandings of legislative history, for instance, but the primary sources presumably are capped by what happened in the past.

133 *See* Samaha, *Tiebreakers*, *supra* note 13, at 1691, 1693 fig.6. In multiple-source situations, holding one source aside does tend to increase the chance of a tie among the remaining sources, but this increase is more than offset by the decisive power of the tiebreaker.

134 *See id.* at 1694.

135 *See id.* at 1714–16. In criminal cases, federal judges leave lenity in a last resort tier below legislative history, and they rule out agency deference. *See id.* In civil cases, federal judges have yet to settle on how to treat legislative history when an agency interpretation is available. *See supra* note 89.

As with decision costs, however, the decisiveness justification for lexical ordering turns on unresolved valuation issues. One question is the frequency of ties in litigation, either objectively or as felt by judges. Judges should sacrifice less to avoid rare events. And sacrifice they must. Lexical ordering obtains gains in decisiveness, and somewhat lower decision costs, at the expense of a more complete picture of the universe of valid sources and their interactions in some fraction of cases.[136] Following our numerical illustration, the probability of an incorrect result is 8% with two sources and one of them a tiebreaker; with five such sources, the probability of such error falls to around 5%—but not zero.[137] Judges can reduce the risk of incorrect case results by carefully selecting weaker sources for the tiebreaking tier, but they cannot eliminate the inaccuracy risks.[138] More work is required to firmly conclude that such trade-offs are justified. We can only add indecision risks as a plausible if supplemental reason for lexically ordering certain types of sources.

* * *

TABLE 1: COMPARING APPROACHES, THEORETICALLY

	Inclusion	*Lexical Ordering*	*Exclusion*
Confronting Trade-offs			
Cost	High	Middling	Low
Quality	High	Middling	Low
Decisiveness	Middling	High	Low
Building Good Hinges	–	Diagnostic, sight unseen	–
Choosing Tiebreakers	–	Costly, shoddy, decisive	–

Note: Weighting sources is required for all three approaches, but less so with fewer sources.

Now we can consolidate several lessons, as a matter of general theory applied to net-positive-value categories of sources. *Excluding sources* will tend to minimize decision costs, but also reduce decision quality and decisiveness. *Including sources* will tend to have the opposite effects, maximizing decision

136 *See* Samaha, *Tiebreakers*, *supra* note 13, at 1694–98 (illustrating erroneous results and other information losses numerically, then introducing the possibility of "double counting" by placing a source in both higher and lower tiers). Overt double counting is a logically promising option, but it is not the focus of contemporary debate or legal practice. Randomization to break ties does not require information losses. *See id.* at 1689–91. But overt randomization often is not available to resolve legal disputes. *See* GUIDO CALABRESI & PHILIP BOBBITT, TRAGIC CHOICES 41–44 (1978) (critiquing lotteries and comparing other allocation rules).

137 *See* Samaha, *Tiebreakers*, *supra* note 13, at 1694 n.86, 1696 fig.8.

138 *See id.* at 1695, 1696 figs.8 & 9 (showing percentages of overlooked ties in or near single digits, and a much higher percentage of cases in which top-tier sources differ from the aggregated implications of all sources).

quality and increasing decisiveness but maximizing decision costs, too. *Lexically ordering sources* will tend toward a unique mixture of effects depending, importantly, on which sources are placed in which tiers and how the doctrinal hinge between tiers is built. If the tiers are sensibly populated and if the hinge quickly and effectively separates cases in which lower-tier sources are less and more likely useful, then lexical ordering will tend toward a middling range of decision costs, a middling range of decision quality, and the highest rate of decisiveness. Loosening the hinge to lower-tier sources will push lexical ordering toward inclusion of those sources; tightening the hinge will push toward exclusion.

C. Convenient Compromise

There is another conceivable defense, or at least explanation, for lexically ordered statutory interpretation. Judges might try to build a united front on broad methodological issues, partly to guide decisionmaking and partly for public relations. Many observers seem to support interpretive methods that reflect orderly decisionmaking with minimal influence from the personal opinions of the judges who happen to sit on cases.[139] Many judges themselves, especially in lower courts, might want sharp guidance for at least getting started on the millrun of statutory cases. And an official method for an entire court system might help observers plan their affairs or draft legislation. At the same time, some number of judges disagree deeply over the proper method and mission for statutory cases. Methodological compromises that are too ambitious will lose support from judges at the extremes—absent effective means of policing how judges actually think—while compromises that are too modest will not guide judges or gain respect from observers who care.

Even with a demand for compromise, we have to wonder why lexical ordering with a clarity test would be part of it, in various fields, times, and jurisdictions. As a start, we may largely rule out a hardline textualism that flatly excludes legislative history and deference to agencies. That option is opposed by too many judges in most jurisdictions to succeed as a global compromise, for now. For now, we also can doubt the widespread viability of an all-inclusive approach with low weights for legislative history and agency interpretations.[140] That option is plausible as a logical matter, and it might

139 *See* Richard A. Posner, *The Jurisprudence of Skepticism*, 86 MICH. L. REV. 827, 887 (1988) (suggesting the vulnerability on this ground of judges using an eclectic mix of open-ended factors); *see also* Micah Schwartzman, Essay, *Judicial Sincerity*, 94 VA. L. REV. 987, 1004, 1006 (2008) (contending that legitimacy requires that officials have reasons that others can understand and accept as a matter of principle, and that judges' reasons should be given publicly); *cf.* Gluck & Posner, *supra* note 16, at 1341–43, 1353 (reporting that several interviewed circuit judges were concerned about "legitimacy" when resolving statutory cases, partly in terms of public support).

140 *See* Jonathan T. Molot, *Ambivalence About Formalism*, 93 VA. L. REV. 1, 51–52 (2007) (suggesting a source-inclusive approach that nonetheless assigns heavy weight to textualist clues).

track judicial practice in many chambers. But something like an all-inclusive approach was tried out as the official line on interpretation in many courts by the 1970s and, arguably, was thought not to produce enough guidance and observer respect. The source weights never reached number-like rigidity, of course. As an alternative, plain meaning rules were familiar and could be grabbed off the doctrinal shelf to support a renewed compromise.[141]

On the other hand, plain meaning rules already had their day in court, too, in the early twentieth century. Guidance was not their strong suit, really, and their connection to rule-of-law values was unproved. The contemporary form of lexically ordered statutory interpretation is likewise. It relies on a hazy clarity test plus multiple sources that still must be assigned weights, if only implicitly.[142] And to the extent that lexical ordering effectively weights lower-tier sources, we should doubt that it has any unique ability to calibrate ex ante incentives for lawyers and legislatures.[143] It tends to moderate the expected influence of lower-tier sources in court, for whatever that's worth, rather than offer an especially instructive backdrop for predicting case results or drafting legislation. Which means that observers need not have been greatly impressed by judicial announcements that lexical ordering was back in fashion. We still have to wonder why many judges would re-endorse lexical ordering for statutory cases by the 1990s.

One possibility is that lexical ordering instructions are more effective than the skeptics suppose, and that judges largely know and tolerate the resulting trade-offs.[144] The decision-cost savings from ignoring sources like legislative history are substantial, and they grew to the extent that legislatures produced more of it per code provision over time.[145] Agency interpretations, where they exist, are highly likely to offer decisive answers to hard interpretive questions. Judges might have learned all this over time. Or if no new lessons were learned, then the mix of methodological commitments on the bench by the 1990s might have returned to about what it was in the 1920s. Furthermore, observers might have picked up on the actual bite of lexical ordering instructions, then adjusted their confidence levels and behavior accordingly. Abbe Gluck's suggestion that lexical ordering likely increases predictability in statutory cases is broadly consistent with these thoughts.[146] Sure, in some cases judges might abandon the compromise in

141 *See, e.g.*, Gluck, *supra* note 12, at 1758–59, 1842–43 (identifying lexical ordering as a viable state-level compromise that is text centric yet not hardline textualism, and that might discipline and legitimize purposivism).

142 *See supra* text accompanying note 89.

143 *See supra* note 128.

144 *See supra* Section II.B.

145 *Cf.* Richard A. Danner, *Justice Jackson's Lament: Historical and Comparative Perspectives on the Availability of Legislative History*, 13 DUKE J. COMP. & INT'L L. 151, 168–70 (2003) (discussing increased production and distribution of congressional committee reports and floor debate transcripts in the late 1800s and thereafter).

146 *See Gluck*, *supra* note 12, at 1856–57 (comparing tiering to open-ended eclecticism). Gluck recognized the presence of debatable empirical questions here. *See id.* at 1858.

favor of other strong commitments,[147] but the identified trade-offs would persist.

Another possibility is that lexical ordering became popular with judges because the instructions have little bite, just as the skeptics suppose, even when judges try to follow the instructions. Judges might then comfortably declare a strong commitment to plain meaning, when they happen to see it, without losing or gaining very much. This could be true for several different reasons. Perhaps the instructions are comprehensive and the clarity test has real meaning to judges, but in operation the test yields results that conveniently fit different judges' methodological commitments. Judges' unadulterated propensity to see clarity might rise along with their principled opposition to lower-tier sources.[148] Or perhaps the instructions are too vacuous to prevent conscious or unconscious manipulation by judges. Judges' tactically inflected propensity to see clarity might depend on their attraction to a given lower-tier source.[149] Or perhaps judges cannot actually ignore lower-tier sources in practice, regardless of their feelings about those sources.[150] None of these depictions seem useful for producing public respect, let alone influencing legislation. But perhaps adequate levels of observer respect are taken for granted or easily sustained.

Some of the above possibilities depend on judges understanding the effects of lexical ordering instructions, while others do not. Yet each one of these positive and normative theories depends on conjecture about judicial behavior. Some theories suppose that lexical ordering has a significant effect on judicial decisionmaking, while others suppose not. And it is not hard to imagine that lexically ordered statutory interpretation became popular again by the 1990s without a thorough evaluation of the consequences for judicial decisionmaking, litigation behavior, and legislative drafting. Nobody fully understands the consequences today, actually.

Fortunately, many of the theories are testable to an extent. We do not have to rest with casual speculation about judges and impressions from their opinions. Instead we need evidence of how lexical ordering actually works.

147 *See id.* at 1757 (suggesting that an agreed-upon methodology might increase fairness, predictability, and coordination by guiding judicial decisions in "less politically charged cases," if not in "hard cases").

148 *See infra* Section III.B. Supreme Court opinions apparently offer an accommodation to justices who oppose the use of legislative history: they are not required to consult it even if the statute is otherwise unclear. *See* Honeycutt v. United States, 137 S. Ct. 1626, 1635 (2017) ("For those who find [legislative history] relevant . . . ."); Milner v. Dep't of Navy, 562 U.S. 562, 574 (2011) ("[F]or those who take it into account . . . ."). *Chevron* doctrine has no such accommodation, thus far, perhaps because many textualists have not yet repudiated deference.

149 *See infra* Section III.C. Consider the position of two textualists, who observed that the plain meaning rule is "essentially sound but largely unhelpful, since determining what is unambiguous is eminently debatable." SCALIA & GARNER, *supra* note 15, at 436.

150 *See infra* Section III.C.

III. Implementation Challenges

Machines can lexically order information without difficulty, we know. Indeed a patch of scholarship on machine learning uses lexical ordering to develop search strategies.[151] But of course people often process information differently from machines, and the specifics of how people process conditionally relevant information are still emerging. The discussion below begins with a review of suggestive prior research, which for the most part does not test lexical ordering instructions or does not test judges, and then reports results from a new survey and vignette experiment. All told, the evidence indicates curiously mixed judicial success at lexical ordering—and therefore a revised picture of the trade-offs.

A. Prior Research

Ordinary people are not always motivated to ignore information and, even if fully motivated, they might not be able to ignore it on purpose.[152] One version of the challenge involves evidence suppression instructions that backfire, making people more likely to rely on information that they have been told to forget.[153] More generally, exposure to some information can influence how people process other information.[154] These phenomena pose implementation challenges when decisionmakers have easy access to sources that are supposed to be reserved for contingencies.[155]

For their part, judges might be in a special class with special abilities.[156] Yet their decision environments are not designed to enhance the success of lexical ordering. Judges have easy access to the lower-tier sources that they

151 *See, e.g.*, Michael Schmitt & Laura Martignon, *On the Complexity of Learning Lexicographic Strategies*, 7 J. Mach. Learning Res. 55, 57–58 (2006).

152 *See* Andrew J. Wistrich et al., *Can Judges Ignore Inadmissible Information?: The Difficulty of Deliberately Disregarding*, 153 U. Pa. L. Rev. 1251, 1260, 1262 (2005).

153 Mock jury research on curative instructions investigates this backfire phenomenon, with mixed findings. *See id.* at 1276 & n.107 (collecting studies). For ironic process theory generally, see Daniel M. Wegner, *Ironic Processes of Mental Control*, 101 Psychol. Rev. 34, 35 (1994).

154 *See* Wistrich et al., *supra* note 152, at 1264–70. A related example appears in first-impressions research where evaluations sometimes are more heavily influenced by early information. *See, e.g.*, Kurt A. Carlson et al., *Leader-Driven Primacy: Using Attribute Order to Affect Consumer Choice*, 32 J. Consumer Res. 513, 513–15 (2006).

155 *See* Wistrich et al., *supra* note 152, at 1275–76 (generalizing that efforts to ignore inadmissible information are more likely to succeed if the information is (1) of questionable credibility, (2) unnecessary to reach a sound decision, (3) not especially salient and not emotionally charged, or (4) not combined with a heavy cognitive load).

156 *See* Frederick Schauer, *Is There a Psychology of Judging?*, *in* The Psychology of Judicial Decision Making 103, 104–05 (David Klein & Gregory Mitchell eds., 2010) (discussing the possibility that judges are unique decisionmakers with regard to legal interpretation); *cf.* Stephan Landsman & Richard F. Rakos, *A Preliminary Inquiry into the Effect of Potentially Biasing Information on Judges and Jurors in Civil Litigation*, 12 Behav. Sci. & L. 113, 125–26 (1994) (concluding that neither judges nor potential jurors were able, when instructed, to ignore evidence of subsequent remedial measures taken by a defendant in a tort case). On

are supposed to bracket mentally, partly because generous rules of advocacy allow lawyers to emphasize basically any interpretive source that they think useful to their cause. We have to wonder what happens to lexical ordering efforts in court, absent special precautions.

A problem area is the "clarity" test, which might be too vague to prevent judges from accessing or blocking lower-tier sources at will. This concern gains some support from vignette experiments conducted by Ward Farnsworth, Dustin Guzior, and Anup Malani.[157] First-year law students were less likely to respond that, in their opinion, criminal statutes were ambiguous when the students' policy preferences were strong.[158] On the other hand, this negative correlation tended to disappear when the question was phrased as whether ordinary readers would likely disagree about the better meaning.[159] Thus a thoughtfully framed clarity question might cheaply reduce the influence of extralegal preferences. Either way, however, cognitive challenges persist for lexical ordering. Judges must ignore the implications of lower-tier sources while working with top-tier sources, whether or not their policy preferences are engaged. The Farnsworth group's study was not designed to examine the influence of lower-tier sources. In addition, as the authors rightly suggest, we should hesitate to draw inferences about judicial behavior from student respondents.[160]

More recently, Dan Kahan and colleagues found that judges and lawyers did respond differently from law students and the general public.[161] The Kahan team drafted two vignettes in which criminal statutes were supposed to be ambiguous, and then manipulated the descriptions of vignette characters in legally irrelevant ways. Thus a littering case involved water dispensers left at the U.S.–Mexico border, either by immigrant aid workers or by border fence construction workers.[162] The judges and lawyers tended to find no violation of the statute regardless of the manipulation, nor did their judgments correlate with their estimated cultural worldviews,[163] but the law students and general public respondents showed greater sensitivity to the

research suggesting that experts are more confident than nonexperts and yet may exhibit various cognitive biases, see DHAMI, *supra* note 124, at 1449–50.

157 *See* Farnsworth et al., *supra* note 17, at 260–62.

158 *See id.* at 266–68. The students apparently were not asked for their final judgments, however. *See id.* at 260–62.

159 *See id.* at 271–72, 292.

160 *See id.* at 272–73.

161 *See* Dan M. Kahan et al., *"Ideology" or "Situation Sense"? An Experimental Investigation of Motivated Reasoning and Professional Judgment*, 164 U. PA. L. REV. 349, 374–75 (2016) (comparing responses from state trial and appellate judges, lawyers in several states, students from five law schools, and a nationally representative sample of the public).

162 *See id.* at 380–83.

163 *See id.* at 394, 396 (showing point estimates); *id.* at 397–402 (showing regressions and Bayesian analysis). The surveys did not ask about the respondents' immigration policy preferences, but the respondents' cultural outlooks were estimated in accord with Cultural Cognition Worldview Scales. *See id.* at 376.

manipulation.[164] The experiment studied legally irrelevant facts rather than lexically inferior sources. However, these facts were ideologically charged and, regardless, the results suggest that judges fall within a special category of legal decisionmakers.

Those results are broadly consistent with a study of treaty interpretation by Yahli Shereshevsky and Tom Noah.[165] Their vignettes asked international law experts and international law students about treaty violations,[166] exposing the treatment groups to treaty preparatory work. The students, but not the experts, seemed inappropriately influenced by the exposure. Isolating those experts who reported that the applicable rule of construction required lexical ordering as well as the exclusion of preparatory work, the authors could not confirm that exposure to preparatory work influenced judgments about treaty violations.[167] As for possible effects of lower-tier sources on ambiguity determinations, Shereshevsky and Noah note that there was no treatment effect in either expert vignette, but that there was a treatment effect in a student vignette involving deportation and terrorism.[168] It is worth pointing out that the experts and students were not given the same vignettes. Also, few of the experts were judges or arbitrators; about two-thirds were academics.[169] Although judges and academics might behave similarly in vignette experiments, we do not have much evidence to go on. That cautionary note aside, the study does suggest that legal professionals can at least sometimes lexically order sources when interpreting legal texts.

In potential contrast, consider a labor-intensive experiment in which judges dealt with an international war crime charge. Holger Spamann and Lars Klöhn asked a group of mostly federal trial-level judges to expend up to fifty minutes reading facts, statutory text, and briefs, with access to the

164 *See id.* at 394, 396–402.

165 *See* Yahli Shereshevsky & Tom Noah, *Does Exposure to Preparatory Work Affect Treaty Interpretation? An Experimental Study on International Law Students and Experts*, 28 EUR. J. INT'L L. 1287 (2017).

166 *See id.* at 1294, 1303 & n.57 (noting that the law students were enrolled in international law courses in Israel and that the experts were recruited from various online Western sources).

167 *See id.* at 1297–1300, 1305–06, 1308–09 (reporting results for "Hypothesis 2"). Their Hypothesis 2 is comparable to our Hypothesis 2, which is described below in subsection III.C.2. Other, more complex departures from lexical ordering instructions are possible and deserve study, as well.

168 *See id.* at 1297 n.38. The student vignette with the treatment effect involved the deportation of a member of Hamas who had helped plan terrorist attacks. *See id.* at 1298. The experts were not given that deportation vignette. The expert vignettes involved the arbitration of a contract (not given to the students) and the detention of a drunk man (given to the students).

169 *See id.* at 1303 (noting that four of 212 experts were international judges or arbitrators); *id.* at 1313 (flagging the possibility that international law experts might be relatively formalistic, in response to concerns that international law is not lawlike).

lengthy trial chamber opinion and an appellate chamber precedent.[170] Variations in the materials were intended to portray the nonbinding precedent as either weakly favoring or weakly opposing the defendant's position on the statute, and to portray the defendant as more or less sympathetic in legally irrelevant ways. The judges' willingness to affirm the conviction was significantly influenced by the legally irrelevant description of the defendant, but not by the difference in the precedent's content.[171] The number of judges in each condition was small,[172] trial judges were asked to behave as appellate judges in an unfamiliar international law context, and lexical ordering was not at issue. But the experimental setting was more realistic than most, and an effect from defendant characteristics was more challenging to find with a small number of respondents. It is not obvious how to reconcile these results with other studies.[173]

Finally, judges achieved mixed success at ignoring inadmissible evidence in experiments by Chris Guthrie, Jeffrey Rachlinski, and Andrew Wistrich.[174] In one experiment, trial judges received a description of a $200 armed robbery, and the treatment group was informed that the defendant confessed during an interrogation that persisted after he asked for a lawyer.[175] The treatment group judges who ruled that the confession was inadmissible were not significantly more willing to convict than were the control group judges.[176] In a follow-up experiment, the Guthrie team added variations in the severity of charges and the police misconduct.[177] This time the pattern of influence was complex. "More severe police misconduct reduced the judges' willingness to convict" compared to mild misconduct, "but only for a less serious crime."[178] When instead the defendant was charged with robbery

170 *See* Holger Spamann & Lars Klöhn, *Justice Is Less Blind, and Less Legalistic, than We Thought: Evidence from an Experiment with Real Judges*, 45 J. LEGAL STUDS. 255, 260–62, 261 n.4 (2016).

171 *See id.* at 268–72; *id.* at 256, 264–65, 273 (emphasizing that the precedent was meant to be only a weak source—either a brief dictum or plausibly distinguishable on its facts).

172 *See id.* at 269 tbl.2 (reporting results with six to eleven judges in each of four conditions).

173 *See id.* at 259 (suggesting greater cognitive load and the "personal implications" of varying the defendant portrayal as possible influences). Perhaps the portrayals of the defendants were rather easy for these judges to understand and process, in contrast to the rest of the subject matter.

174 *See* Wistrich et al., *supra* note 152, at 1258–59, 1286–1312.

175 *See id.* at 1320–21, 1344–45.

176 *See id.* at 1321; *see also id.* at 1316, 1342–43 (reporting results of an experiment in which exposure to the fruits of a car search did not significantly affect the judges' determinations of probable cause).

177 *See* Jeffrey J. Rachlinski et al., *Altering Attention in Adjudication*, 60 UCLA L. REV. 1586, 1611–13 (2013) (noting that the bank robbery involved $520 and that some judges were told about an ensuing murder of a young mother; and that judges were either not told about a confession, told about a written confession after the police twice refused the defendant's request for a lawyer, or told about the same confession after nine hours of police threats and other misconduct).

178 *Id.* at 1614.

plus murder, "the judges who had heard confessions, however obtained, were consistently more willing to convict."[179]

Perhaps lexically ordered statutory interpretation is easier for judges to handle. The inadmissible evidence experiments include emotionally charged material of a kind not present in the ordinary lower-tier source for statutory cases. But we have awfully little on-point evidence. To understand how lexical ordering operates in our judiciaries, we need a better grip on how judges use clarity tests and how lower-tier sources influence interpretation.

B. Clarity Survey

On that score, many commentators have doubted that clarity tests can have much effect on judges.[180] One version of the idea is that a clarity test will have little bite across different methodological commitments. Even if judges do ignore lower-tier sources when applying the clarity test, each judge might reach a happy ending on the legal relevance of those sources if the tendency to find statutes clear is correlated with the tendency to find lower-tier sources useless. Judges who are skeptical (enthusiastic) about legislative history might usually conclude that statutory text is otherwise clear (unclear). This pattern could hold even if judges engage in no case-specific manipulation of the test.

A simple way to gain evidence on this point is to ask judges about their experiences. Approximately one hundred appellate judges from around the country were presented with general questions about text and legislative history.[181] Each judge was asked, "based on your experience deciding statutory cases, how often do you end up thinking that the statute's text is ambiguous or vague," on a scale of 1 to 7 (1 indicated never and 7 indicated always). Each judge also was asked, "based on your experience deciding statutory cases that involve legislative history, how often is the legislative history useful to you," on the same kind of scale. If a judge's ability or willingness to find textual clarity is associated with the judge's lack of attraction to legislative history, then we might see the numerical responses on these questions run-

179 *Id.* For another of the few studies on expert ability to ignore information, consider the fascinating treatment of fingerprint analysis in Itiel E. Dror et al., *Contextual Information Renders Experts Vulnerable to Making Erroneous Identifications*, 156 FORENSIC SCI. INT'L 74, 76 (2006). The study took place in the experts' workplace, with collaborators asking experts to give an opinion on a terrorism case that had been adjudicated. One set of collaborators told their experts that a match had been confirmed in the terrorism case. In truth, the experts were not given fingerprints from the terrorism case, but instead given fingerprints that the experts themselves had analyzed in a real prior case and had concluded did not match. Most of the experts who were given the promatch indication changed their judgment to match or unclear. The number of experts in the study was small, but understandably so given the impressive research design.

180 *See supra* notes 15–17 (collecting sources).

181 For additional details on the survey, see subsection III.C.1, which notes that most of the judge respondents were state judges who had served for four or fewer years in their current positions.

ning together and perhaps clumping toward the high and low ends of the scales.

The results provide only soft support for the suggested pattern. Figure 1 is a scatterplot of responses.[182] Some judges did indicate that statutory text is usually vague while legislative history is usually useful, and a larger group gave the opposite combination of responses. But those groups did not dominate. The combined fraction of survey responses that were below four for both questions or above four for both questions comprised less than one-third of the sample. While the responses on textual vagueness clustered at three, four, and five, the responses on legislative history's usefulness were fairly spread out. The correlation between the numerical responses on the two questions is weak ($r = 0.09$) and not statistically significant at conventional levels ($p = 0.38$). Surely some judges have general tendencies that make the clarity test insignificant to them, but the size of this carefree group is open to question and it might well be small. Which would tend to preserve the hard trade-offs identified in Section II.B.

These measures are imperfect, of course. The responses are self-reported generalizations at one point in time, and not every judge necessarily had in mind the same conception of legislative history.[183] Furthermore, a judge's estimate of legislative history's usefulness in general could be influenced by recent exposure to a concrete example. Our treatment group of judges was given a reference to committee reports,[184] which have become controversial, and the reports were in tension with some other sources. This could pull down the treatment group's estimate of legislative history's usefulness.[185] The treatment group's mean for the usefulness question (3.71) was about one-half point lower than the control group's mean (4.24). However, this difference is not statistically significant at conventional levels ($p = 0.14$), and an ordered probit analysis fails to confirm a treatment effect on the usefulness response ($p = 0.238$). Exposure to committee reports probably did not have a large effect on these responses.

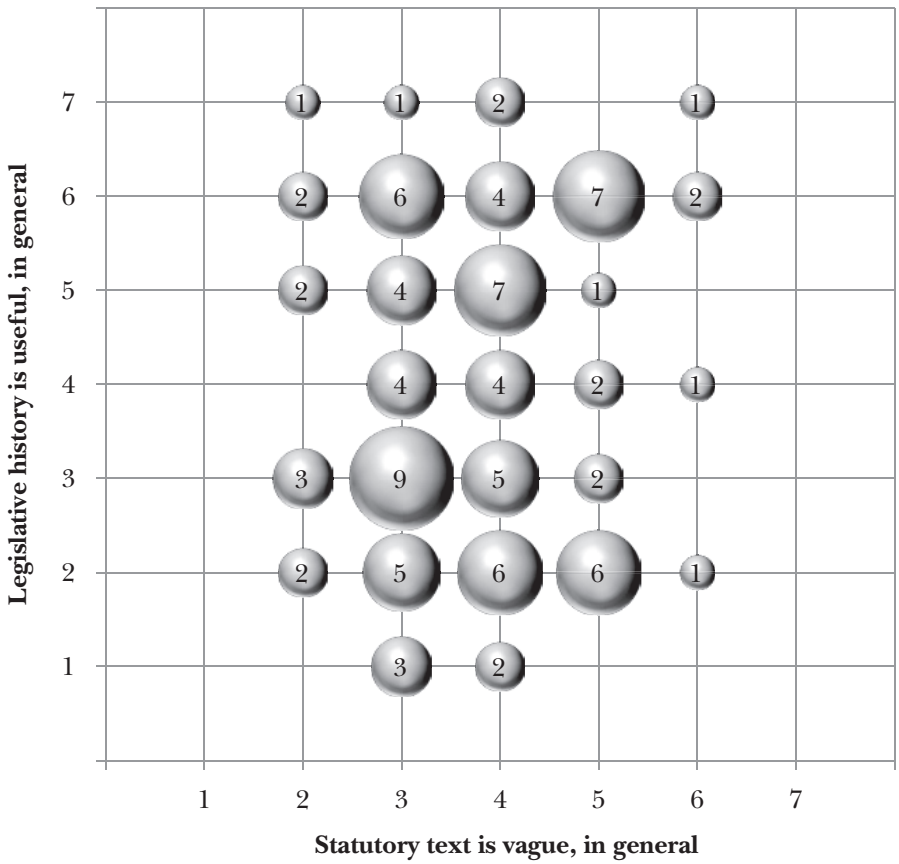
182 One judge wrote "3–4" in response to both questions. That survey is excluded from Figure 1.

183 *See supra* subsection I.B.1 (observing that available state-level legislative history varies). Perhaps some judges who reported that legislative history is not usually useful to them are trying to tell us that they are blocked from making legislative history useful *because they follow a plain meaning rule* and usually conclude that the text is clear. I attempted to avoid this problem with the wording of the question, which asks about "cases that involve legislative history."

184 *See infra* subsection III.C.1 (detailing the vignettes).

185 It is not obvious how such a downward pull would affect the correlation between responses on the legislative history and statutory clarity questions. As it happens, this correlation looks weaker in the treatment group ($r = -0.03$, $p = 0.83$) than in the control group ($r = 0.19$, $p = 0.20$), but neither correlation is statistically significant at conventional levels.

FIGURE 1: Scatterplot of appellate judge survey responses, 2016–2017 (n = 95)



C. *Vignette Experiment*

Our survey collected some general impressions from judges. We need another approach to better understand the interplay between tiers of sources. Below are the results from a vignette experiment that begins to measure judicial sensitivity to particular lower-tier sources in particular cases.

1. Participants and Procedure

All of the participants were sitting appellate judges. Of the 102 respondents, eighty-six were attendees at two seminars for new appellate judges from around the country,[186] and another sixteen were intermediate appellate judges in a Midwestern state who attended an in-house educational con-

186 The Institute of Judicial Administration's New Appellate Judges Seminar is held annually in the summer. The experiment was conducted at the 2016 and 2017 Seminars. A different group of student judges attends each year.

ference. Most respondents were serving on a state court (seventy-eight), others were federal circuit court judges (eight) and military judges (ten), and a few respondents left this question unanswered (six). Most of the judges in the pool had fewer than five years of experience in their current positions, which was a requirement for student judges attending the two seminars.[187] However, the majority of the student judge attendees had prior judicial experience of another kind, such as in a trial court,[188] and nearly all judges in the Midwestern judges group had served for at least five years on that court. We cannot identify the experience level of each respondent, though. To protect anonymity and to increase response rates, the survey did not include demographic questions.

Each judge received two hypothetical cases to decide by paper and pencil survey. The vignettes are reproduced in Appendices A and B. In *Trade Name*, one company claims that another company is unlawfully using the same name for a different drug. The plaintiff company is not yet selling its drug on the market, however, and the disputed issue is whether presale safety testing counts as being "used in commerce openly" under a quoted statute.[189] In *Election Law*, plaintiff employees who were paid an hourly wage claim that their employer unlawfully refused to pay them for the time that they left work to go vote. The disputed issue is whether such refusal to pay counts as "penalizing . . . with a reduction in wages" under a quoted statute.[190]

187 *See* INST. OF JUDICIAL ADMIN., *New Appellate Judges Seminar*, N.Y.U. L., http://www.law.nyu.edu/centers/judicial/appellatejudgesseminar (last visited August 25, 2018).

188 *See New Appellate Judges Seminar: 2016 Attendees*, IJA REPORT (The Inst. of Judicial Admin., New York, N.Y.), Fall 2016, at 11–22, https://issuu.com/nyuija/docs/final_ija_newsletter_highres (showing that fifteen of twenty-nine nonmilitary confirmed student judge attendees had prior judicial experience); *New Appellate Judges Seminar: July 16–27, 2017 Attendees*, IJA REPORT (The Inst. of Judicial Admin., New York, N.Y.), Winter 2017, at 6–21, https://issuu.com/nyuija/docs/ija_report_winter_2017_final_single (showing that twenty-five of forty-two nonmilitary attendees had such prior experience). Faculty judges have more experience than the student judges. At the 2016 Seminar, both student and faculty judges were allowed to participate to keep the number of respondents up. Because several of the same faculty judges attended the 2017 Seminar, these faculty judges were excluded from the pool.

189 I use "trade name" loosely; this vignette is a variation on a federal trademark law issue. *Cf.* 15 U.S.C. § 1127 (2012) (defining "use in commerce"); Kythera Biopharmaceuticals, Inc. v. Lithera, Inc., 998 F. Supp. 2d 890, 899 (C.D. Cal. 2014) ("Shipments of drugs for clinical testing may be a sufficient use in commerce to show a protectable interest."); S. REP. NO. 100-515, at 44–45 (1988), *as reprinted in* 1988 U.S.C.C.A.N. 5577, 5607 (stating that the definition should be interpreted flexibly to include less traditional uses such as "ongoing shipments of a new drug to clinical investigators"). This issue seems fairly obscure for appellate judges, especially state judges.

190 This vignette is a variation on pay-while-voting disputes from the mid-twentieth century. *Cf.* Day-Brite Lighting, Inc. v. Missouri, 342 U.S. 421, 422 n.1, 425 (1952) (rejecting a constitutional challenge to a state statute that expressly referred to a "penalty" or "deduction of wages"); State v. Int'l Harvester Co., 63 N.W.2d 547, 550 (Minn. 1954) (similar); People v. Ford Motor Co., 63 N.Y.S.2d 697, 697 (N.Y. App. Div. 1946) (per curiam) (similar).

Both vignettes were organized similarly, with a short description of the disputing parties, statutory text quoted, and party arguments back and forth. Both vignettes include arguments about ordinary meaning based on dictionaries, meaning within a larger textual context, and consistent use across statutory texts in light of judicial precedent. The vignettes in the control condition were intended to moderately favor the defendants as a matter of conventional legal analysis, although judge responses to the Election Law case were difficult to predict. That vignette was intended to implicate a more ideologically charged field and was inspired by prior observational studies of judge voting behavior.[191]

A randomly assigned treatment group received the same descriptions as above, plus additional information and lexical ordering instructions. In Trade Name, the treatment group (n = 51) was given on-point legislative history, while the control group (n = 50) was not. The treatment group was told that uncontradicted committee reports indicate that the statutory language was intended to include safety testing. The message from this source clearly favored the plaintiff-company's position, although judges were not instructed on how much weight to give this source. In Election Law, the treatment group (n = 50) was given an administrative agency interpretation that clearly favored the plaintiff-employees' position, while the control group (n = 51) was not. The agency's position included a short explanation of the statute's purported intent, with reference to employees having the same freedom to go to the polls as business owners. The treatment group was told that the courts in their jurisdiction have decided that judges should consider committee reports and should defer to an agency's position "if but only if the other sources of statutory interpretation, taken together, leave the statute's meaning fairly or totally unclear as applied to the facts of the case."[192]

All judges were asked about statutory clarity at the end of each vignette. The control group was asked, for each vignette, "how clear is the statute as applied to the facts of this case, on a scale of 1 to 5?"—with notations briefly describing the scale (1 = totally unclear, 2 = fairly unclear, 3 = somewhat clear, 4 = fairly clear, 5 = totally clear).[193] These questions are doctrinally extraneous in the absence of lexical ordering and lower-tier sources, but the questions facilitate direct comparisons with the treatment group. The treatment group's parallel questions reiterated the lexical ordering instructions

---

Respondents might have been aware of their home-state law on the subject. This awareness could influence those judges' sense of statutory clarity in the vignette and the correct judgment. But it is unclear how such awareness would affect the influence of the hypothetical *lower-tier source* presented to the treatment group. One possibility is that awareness of home-state law would tend to reduce the influence of sources in the vignettes, because the answer became foreordained, thereby biasing the resulting treatment effect downward.

191 *See* Samaha, *Looking Over a Crowd*, *supra* note 117, at 590, 610–13 (discussing the plausibility and evidence of ideological influence in voting rights cases and in trademark cases).

192 *See infra* Appendices A & B.

193 *See infra* Appendices A & B.

by asking, "how clear is the statute as applied to the facts of this case, without considering [the committee reports/the agency's position], on a scale of 1 to 5?"[194] The same descriptive notations were given. All judges were then asked for a final judgment in each case and for a level of confidence about those judgments.

The survey instrument thus allowed judges to make decisions about statutory clarity and, as a consequence, the legal relevance of lower-tier sources. The clarity scale was described such that a response of three or above meant that the judge was not supposed to consider the given lower-tier source. Because the control group saw no lower-tier sources in their vignettes, we can use their responses as a baseline for testing the effect, if any, of exposing judges to such information.

2. Hypotheses

Following the lexical ordering instructions required the treatment group to respond to the statutory clarity questions while ignoring lower-tier sources. For one or more reasons, however, some respondents might be unable or unwilling to mentally bracket those sources.[195] Exactly how exposure to lower-tier sources might influence clarity responses is difficult to predict, but a significant difference at any point on our five-point scale would suggest a problem for lexical ordering. The primary and simplest hypothesis for a lexical ordering failure is that clarity responses will differ between treatment and control groups ("Hypothesis 1—Clarity Effect").

Following the lexical ordering instructions also required that final judgments be stable, in part. Judges who assigned clarity levels below three were supposed to consider or follow the lower-tier source on final judgment, but not judges who assigned clarity levels of three and above. We can look for evidence that exposure to a lower-tier source influenced certain combinations of judgments and clarity levels, depending on the distribution of responses in the control group.

One possibility is that lower-tier sources will improperly influence final judgments without influencing clarity levels. The lexical ordering instructions would be violated if a judge responded that a statute was at least somewhat clear regardless of exposure to lower-tier sources, but nonetheless switched final judgments because of the exposure. Thus another hypothesis is that, among judges who assign clarity levels three and above, the fraction of judges entering proplaintiff judgments will differ between treatment and control groups ("Hypothesis 2—Judgment Flipping"). This effect would be especially strong, however, and we might have difficulty distinguishing such judgment flipping from more complex failures of lexical ordering.[196]

194 *See infra* Appendices A & B.

195 *See supra* notes 153–57 (collecting explanations for failure to ignore information).

196 Exposure to lower-tier sources might have two simultaneous effects: (1) some judges lower their clarity levels to below three and flip judgments from defendant to plaintiff; (2) other judges stick with proplaintiff judgments but raise their clarity levels to three and

In addition, we can begin to examine narrower theories for the convenient use of clarity tests.[197] Perhaps judges who are attracted to a lower-tier source will make the source legally relevant by migrating to clarity levels below three. Judges who are instead repelled by a lower-tier source might make the source legally irrelevant by migrating to clarity levels three and above. In our vignettes, some fraction of the control group will enter judgment for the defendants, while the treatment group will see lower-tier sources that support the plaintiffs. We can look for evidence in support of the hypotheses that the treatment group will move away from prodefendant judgments at clarity levels three and above, and toward proplaintiff judgments at clarity levels below three ("Hypothesis 3—Attraction"); and/or that the treatment group will move away from prodefendant judgments at clarity levels below three, and toward prodefendant judgments at clarity levels three and above ("Hypothesis 4—Backlash").

Identifying these convenient-use patterns might be difficult. We lack data on exactly how the judges feel about the lower-tier sources, aside from their estimates of legislative history's usefulness in general. Furthermore, the attraction and backlash theories are not mutually exclusive within a group of judges. These theories predict opposite effects on prodefendant judgments at clarity levels three and above, and those effects could cancel out. Fortunately, the attraction theory predicts an increase in proplaintiff judgments at clarity levels below three while the backlash theory does not. However, such an increase could come at the expense of prodefendant judgments at clarity levels below three, which would not indicate a failure to lexically order. Remember that judges who think that the statute is unclear are supposed to consider the lower-tier sources, which support the plaintiffs. In any event, we should entertain plausible alternative explanations for apparent stability or movement in judge responses.

As for differences between vignettes, reasonable debate is possible. One plausible prediction is that, in Trade Name, most judges will be able to follow the instructions and ignore the committee reports in assessing clarity and in making final judgments when they believe that the statute is otherwise adequately clear. Although specific and uncontradicted support for one side in legislative history is a tempting source for some judges, the arguably low policy stakes might help legal professionals execute the instructions. The most plausible prediction for Election Law seems cloudier. The ideological charge of the issue area could reduce the influence of instructions and conventional legal analysis; if so, we might see little difference when judges are exposed to an agency position. But judges who are sympathetic to the plaintiff's position as a matter of policy, constitutional principle, or otherwise might be tempted

above. This combination might be confused with judgment flipping at clarity levels three and above. Also note that crosscutting judgment-flipping effects could cancel out, if some judges are highly attracted to and other judges backlash away from lower-tier sources. Indeed, all of the foregoing lexical ordering failures could occur without leaving clear footprints in our data.

197 *See supra* Section II.C.

to rely on agency support, thereby combining law's sources with extralegal moves.

3. Results

The results are provocatively mixed. We see warning signs that some judges departed from the lexical ordering instruction in the Election Law case, but no equally consistent signs in the Trade Name case. Table 2 presents summary statistics.[198]

TABLE 2: SUMMARY STATISTICS

	Control	Treatment	Difference
Trade Name, means			
Clarity of the statute	3.14 (n = 49)	3.35 (n = 51)	0.21
Judgment for defendant	0.88 (n = 50)	0.71 (n = 51)	-0.18
Confidence in judgment	5.12 (n = 50)	4.69 (n = 46)	-0.43
Election Law, means			
Clarity of the statute	2.94 (n = 51)	3.50 (n = 49)	0.56
Judgment for defendant	0.39 (n = 49)	0.18 (n = 49)	-0.21
Confidence in judgment	4.83 (n = 51)	5.18 (n = 48)	0.34

- Hypothesis 1—Clarity Effect

The first question is whether exposure to lower-tier sources influenced judges' clarity responses on our five-point scale. A Mann-Whitney U test is one way to compare distributions of discrete ordinal data.[199] Using that test

198 We will generally report p-values to the third decimal place, and readers may draw their own lines. *See, e.g.*, ROBERT M. LAWLESS ET AL., EMPIRICAL METHODS IN LAW 233–34 (2010) (noting the $p < 0.05$ convention for statistical significance); Adam S. Chilton & Marin K. Levy, *Challenging the Randomness of Panel Assignment in the Federal Courts of Appeals*, 101 CORNELL L. REV. 1, 36 (2015) (using $p < 0.10$); Philip J. Cook & Jens Ludwig, *Aiming for Evidence-Based Gun Policy*, 25 J. POL'Y ANALYSIS & MGMT. 691, 694 (2006) (contending that the standard of significance for policy-relevant evidence should account for information availability and expected costs and benefits).

199 *See* LAWLESS ET AL., *supra* note 198, at 271–73 (explaining that the test does not depend on a normal distribution, does not rely on variables' magnitudes, and is statistically identical to the Wilcoxon rank-sum test); Farnsworth et al., *supra* note 17, at 267 (using the Wilcoxon rank-sum test for a four-value scale of statutory ambiguity).

Using t-tests to compare means is controversial here because the data are not continuous and normally distributed. For readers willing to consider that comparison: in Trade

in Trade Name, the difference in clarity distributions for the treatment and control groups is not statistically significant at conventional levels ($p = 0.196$). The treatment judges might well have ignored the committee reports while assessing statutory clarity in that case. In Election Law, however, the difference in clarity distributions is highly significant ($p = 0.004$), indicating that exposure to the agency's position influenced some of the treatment judges. We also conducted a Kolmogorov-Smirnov test, which evaluates the difference in the largest gap between points on two distributions.[200] Under that test the result is more equivocal for Trade Name, although the difference between treatment and control does fall short of the 95% threshold ($p = 0.090$). The result for Election Law remains unequivocal, however. The difference in clarity distributions is again highly significant ($p = 0.005$).

We explored the robustness of these results by adding control variables. Tables 3 and 4 report estimates from ordered probit models, and the headlines are broadly similar.[201] Table 3 shows that we cannot confirm a treatment effect in any of our models for Trade Name. Model 1 has one independent variable for whether the judge was in the treatment group. Model 2 adds a variable for the judge's view on the usefulness of legislative history: the Trade Name treatment group was exposed to one form of legislative history, and perhaps a judge's general view on its usefulness will indicate a propensity to find these statutory texts unclear. Model 3 adds an interaction term to check whether any treatment effects are conditional on judges' general views about the usefulness of legislative history.[202] Model 4 adds a variable representing the judge's general view on the vagueness of statutory text. Model 5 adds controls for whether the judge was part of the Midwest state group, a federal judge, or a military judge. Neither treatment nor any other independent variable is significant in any of these models.[203]

Name the difference between the control and treatment means (3.14 versus 3.35) is not significant ($p = 0.269$), while in Election Law the difference (2.94 versus 3.50) is significant ($p = 0.006$). This is largely consistent with our other results.

200 *See* GREGORY W. CORDER & DALE I. FOREMAN, NONPARAMETRIC STATISTICS 80 (2d ed. 2014).

201 *See generally* LAWLESS ET AL., *supra* note 198, at 345, 349–50 (discussing ordered probit models). Our randomized trial already provides some comfort in drawing causal inferences.

202 When interpreting interaction effects in nonlinear models such as ordered probit, it is insufficient to assess the sign and *p*-value on the coefficient for the interaction term. More thorough interpretation requires an assessment of the sign and *p*-value for each possible combination of values of the interaction term. *See* Chunrong Ai & Edward C. Norton, *Interaction Terms in Logit and Probit Models*, 80 ECON. LETTERS 123 (2003). Doing so for Models 3–5 in Trade Name indicates that the signs and *p*-values for each value of the interaction term are all negative and not statistically significant at even the 90% level.

203 In two models not displayed in Table 3, we instead used an interaction term with a dummy variable for whether the judge's view on legislative history's usefulness was especially high on the seven-point scale (i.e., Treatment * LH useful > 5), or especially low (i.e., Treatment * LH useful < 3). None of the independent variables reached statistical significance at the 90% level in Trade Name. We also ran a simpler model with no interaction term but with Treatment, LH useful, and Text vague; then a model with those variables

Table 4 shows estimates for Election Law. The results are somewhat mixed using the models above: The treatment variable is highly significant in Models 1 and 2 ($p < 0.01$) but falls short of the 95% level in Models 3, 4, and 5 ($p < 0.10$). However, those last three models include the interaction term, which involves legislative history's usefulness. Election Law's treatment group was exposed to an agency position, not legislative history, and thus we lose any straightforward reason to include a variable that represents judicial views on legislative history.[204] In Model 6, we drop the variables involving legislative history and include treatment and the textual vagueness variable. In Model 7, we add back the controls for judge type. In these last two models for Election Law, the treatment variable is, once more, highly significant ($p < 0.01$).[205] For the sake of symmetry, we ran Models 6 and 7 for Trade Name and, like the other models for Trade Name, the treatment variable is not statistically significant.

Finally, we conducted a binomial logit regression for which we converted the five-point clarity scale into a binary variable: low (1 or 2) or high (3, 4, or 5). This split tracks the legal line between unclear and clear statutes that was drawn by the lexical ordering instructions. In Trade Name, the treatment variable is again not significant ($p = 0.875$). In Election Law, this time the treatment variable is only marginally significant ($p = 0.096$). This muted result reflects that the differences between the treatment and control groups' clarity responses are mostly at clarity levels three and above, rather than across the legal line between clear and unclear statutes. The distributions of clarity responses are displayed in Figures 2 and 3. Figures 4 and 5 display histograms that combine judges' clarity responses with their corresponding final judgments, which are relevant to our other hypotheses.

plus the controls for Midwest group, federal judge, and military judge; then two models with all of the foregoing variables but swapping out LH useful for either the dummy variable, LH useful > 5, or instead, LH useful < 3. Again, none of the independent variables reached statistical significance in these models for Trade Name.

204 We lack data on whether judges' general views about the usefulness of legislative history track their general views about the persuasiveness of agency interpretations.

205 Additionally, when any of the models discussed in note 203 above are run for Election Law, the treatment variable is significant at $p < 0.01$ or $p < 0.05$.

TABLE 3: TRADE NAME—ORDERED PROBIT MODELS OF CLARITY RESPONSES

Independent Variables	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Treatment	0.223	0.258	0.663	0.670	0.686	0.281	0.335
	(0.215)	(0.227)	(0.572)	(0.575)	(0.578)	(0.225)	(0.230)
LH useful		-0.0119	0.0382	0.0378	0.0385		
		(0.0662)	(0.0927)	(0.0928)	(0.0932)		
*Treatment * LH useful*			-0.102	-0.104	-0.0944		
			(0.132)	(0.133)	(0.134)		
Text vague				0.0137	0.0502	-0.00218	0.0370
				(0.110)	(0.116)	(0.108)	(0.114)
Midwest group					0.459		0.442
					(0.326)		(0.324)
Federal judge					0.353		0.345
					(0.425)		(0.424)
Military judge					-0.196		-0.243
					(0.387)		(0.384)
Observations	101	95	95	95	95	96	96

TABLE 4: ELECTION LAW—ORDERED PROBIT MODELS OF CLARITY RESPONSES

Independent Variables	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Treatment	0.579***	0.586***	1.022*	0.999*	0.992*	0.597***	0.652***
	(0.217)	(0.227)	(0.567)	(0.571)	(0.574)	(0.223)	(0.229)
LH useful		-0.0510	0.00111	0.000404	0.00201		
		(0.0645)	(0.0894)	(0.0895)	(0.0899)		
*Treatment * LH useful*			-0.109	-0.103	-0.0872		
			(0.129)	(0.130)	(0.131)		
Text vague				-0.0344	0.0102	-0.0477	0.00379
				(0.108)	(0.113)	(0.106)	(0.112)
Midwest group					0.163		0.188
					(0.319)		(0.318)
Federal judge					-0.0960		-0.104
					(0.412)		(0.410)
Military judge					-0.577		-0.603
					(0.380)		(0.377)
Observations	101	96	96	96	96	97	97

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. Standard errors in parentheses. The dependent variable is the judge's assessment of statutory clarity on a five-point scale. The Trade Name treatment group was exposed to legislative history; the Election Law treatment group was exposed to an agency position. "LH useful" and "Text vague" are the judge's responses to general questions about the usefulness of legislative history and the vagueness of statutory text, respectively, measured on seven-point scales of frequency. On the interaction term, see *supra* note 202.

FIGURE 2: Trade Name—Distribution of clarity responses

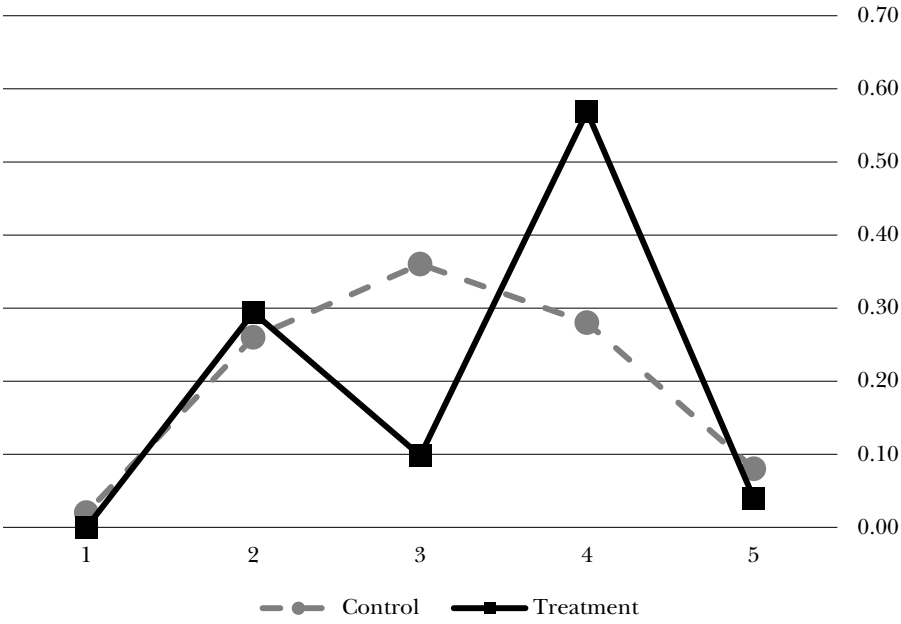


FIGURE 3: Election Law—Distribution of clarity responses

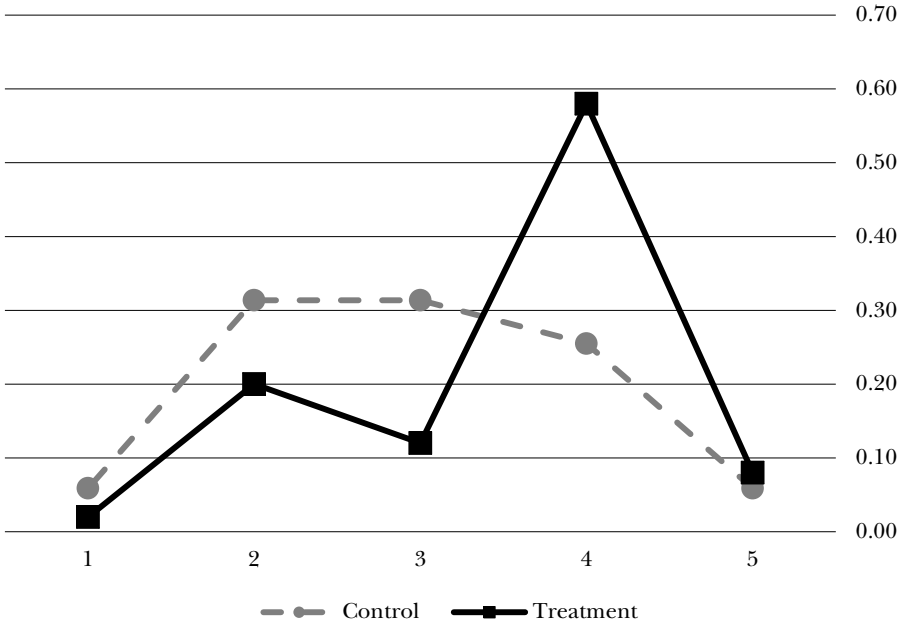


FIGURE 4: Trade Name—Distribution of clarity-plus-judgment responses

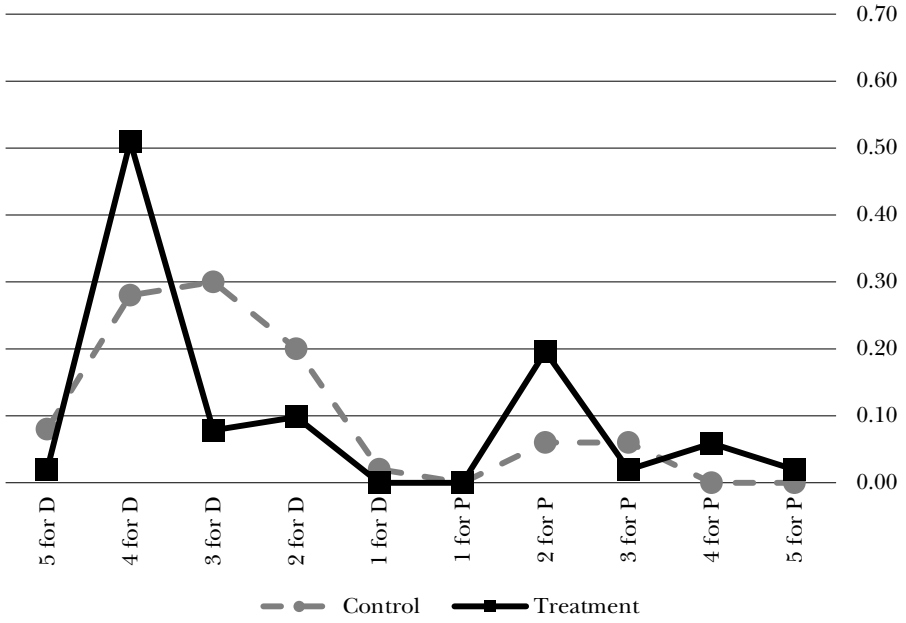
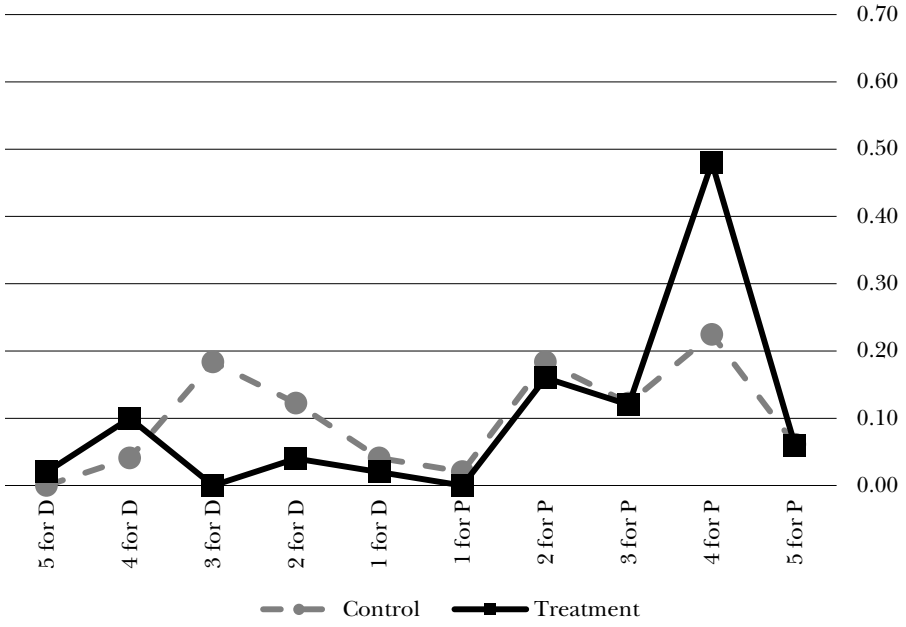


FIGURE 5: Election Law—Distribution of clarity-plus-judgment responses



- Hypothesis 2—Judgment Flipping

To test our hypothesis regarding judgment flipping at high-clarity levels, we isolated the judges who assigned clarity levels of three and above, which were sufficiently high to rule out the lawful consideration of lower-tier sources. In Trade Name, this left seventy-two judges (thirty-six in control, thirty-six in treatment); in Election Law, this left seventy judges (thirty-one in control, thirty-nine in treatment). We then compared the percentage of these judges who entered judgment for the defendant in the treatment as opposed to the control group. Finding a significant difference on this measure would intimate a rather serious lexical ordering failure, in which final judgments are flipped by exposure to a lower-tier source despite consistently high reported levels of statutory clarity in the other sources. This effect would be something like jumping from one end of the histogram to the other.

The Trade Name results do not indicate such judgment flipping. The gap in the percentage of judges who entered judgment for the defendant at clarity levels three and above in the control group (92%) and the treatment group (86%) is not significant ($p = 0.460$). This result is hard to explain on the theory that such excerpts from committee reports are not persuasive enough to influence judgments. Among the relatively few judges who assigned clarity levels below three, the difference in prodefendant judgments in control (79%) and treatment (33%) is significant ($p = 0.013$).

The Election Law numbers are more suggestive of judgment flipping, with our test result hovering around conventional significance thresholds. The difference between prodefendant judgments at clarity levels three and above in control (35%) and treatment (15%) approaches the 95% level in a two-sided test ($p = 0.052$).[206] This is true even though the control group in Election Law already tilted toward the plaintiffs, which put a ceiling on the attractive effect of the proplaintiff agency position. We have reason to suspect that the agency position moved some judges to the plaintiffs' side despite the lexical ordering instructions, perhaps by flipping judgments at relatively high clarity levels. Other, more subtle lexical ordering failures are possible as well.[207]

---

206 Two judges' responses to the final judgment question for the Election Law vignette were difficult to interpret and were excluded from our tests for Hypotheses 2, 3, and 4. One of these two surveys was implicated in the Election Law test for Hypothesis 2 and thus excluded. If instead this response is coded as proplaintiff, the significance of the difference decreases ($p = 0.064$); if instead this response is coded as prodefendant, the significance increases ($p = 0.033$). All three results are near the 95% level.

207 For reasons noted above, *see supra* note 196, our test for Hypothesis 2 cannot readily distinguish between judgment flipping without changes in clarity responses and more complex movements along a continuum. Those more complex movements represent a different type of lexical ordering failure, not a success. In Election Law, the percentage of judges assigning clarity levels of three and above was not obviously the same in control (63%) and treatment (78%)—which is a reason to suspect that more than judgment flip-

- Hypothesis 3—Attraction

As for convenient use of clarity tests, we find little evidence for the specific versions of this idea in our hypotheses. According to the attraction theory, judges who would otherwise enter prodefendant judgments at clarity levels three and above will switch to the plaintiff side and make the lower-tier source legally relevant by moving to clarity levels below three. But we cannot confirm this pattern in our results. In Trade Name, the fraction of judges entering prodefendant judgments at clarity levels three and above did not drop significantly in the treatment group ($p = 0.271$, one-sided t-test), even though the fraction of proplaintiff judgments at clarity levels below three did rise ($p = 0.019$, one-sided t-test).[208] For that latter effect, we cannot rule out a perfectly lawful explanation: The rise in proplaintiff judgments at low clarity levels could have come from treatment judges who otherwise would have entered prodefendant judgments at equally low clarity levels. As long as their clarity assessments were not affected, judges were supposed to consider the proplaintiff committee reports at low clarity levels.

Nor can we confirm this attraction theory in Election Law. Although the fraction of prodefendant judgments at clarity levels three and above might be lower in the treatment group ($p = 0.079$, one-sided t-test), the fraction of proplaintiff judgments at clarity levels below three is not higher ($p = 0.728$, one-sided t-test). Perhaps exposure to the agency position had more than one effect and this attraction theory captures one of them, but we cannot be sure.[209] Either way, and by the way, there is no lawful explanation for any real decrease in prodefendant judgments at relatively high clarity levels. The instructions required judges at those clarity levels to lock in their clarity responses and final judgments against influence from lower-tier sources.

- Hypothesis 4—Backlash

According to the backlash theory, when exposed to the proplaintiff lower-tier sources, judges who would otherwise enter prodefendant judgments at clarity levels below three will stick with the defendant and make the source legally irrelevant by moving to clarity levels three and above. But we cannot confirm this pattern, either. In Trade Name, there might be a drop in prodefendant judgments at clarity levels below three in the treatment group ($p = 0.052$, one-sided t-test), but there is no accompanying increase in

ping occurred. In Trade Name, that percentage was very similar in control (72%) and treatment (71%).

208 One-sided tests are appropriate here because the hypothesis predicts the direction (i.e., the sign) of the two treatment effects.

209 For example, (1) some judges in the treatment group could be responding to the agency's position by moving from prodefendant judgments at clarity level three to proplaintiff judgments at clarity level two, while (2) other judges in the treatment group could be responding by moving from proplaintiff judgments at clarity level two to proplaintiff judgments at clarity level three or four. The combined effects on proplaintiff judgments at clarity level two could cancel out such that no change is noticeable in that slot on our histogram.

prodefendant judgments at clarity levels three and above ($p = 0.729$, one-sided t-test). In Election Law, the upshot is the same. Even if prodefendant judgments at low clarity levels are significantly lower in the treatment group ($p = 0.048$, one-sided t-test), there is again no accompanying increase in prodefendant judgments at clarity levels and above ($p = 0.921$, one-sided t-test).

- Other Possibilities

Several of our hypotheses focused on the legally critical line for statutory clarity. But, in retrospect, Figures 2–5 suggest other effects that do not depend on judges crossing that line. We can pause to think about possible effects that were not predicted by our initial hypotheses and that might warrant future study. The most striking possibility is that some judges reacted to the lower-tier sources by elevating their clarity assessments, to create an extra cushion against the legal line for relying on those sources.

The raw clarity distributions suggest that treatment group judges moved toward clarity level four—not toward law's borderline between clear and unclear statutes. But these apparent spikes favored different sides in the two different cases. Having seen the proplaintiff committee reports in Trade Name, some treatment judges might have backed away from the implication of that lower-tier source and entered prodefendant judgments at clarity level four. In contrast, having seen the proplaintiff agency position in Election Law, some treatment judges might have leaned toward the implication of that lower-tier source and entered proplaintiff judgments at clarity level four.

We cannot be sure where these possible movements would have come from. But few control group judges assigned clarity level five in either case, so the bulk of any movement would have come from lower clarity levels. The largest percentage drop from control to treatment for any clarity-plus-judgment combination is the same in both cases: prodefendant judgments at clarity level three. So, in Trade Name, some treatment judges might have moved from borderline clarity to higher clarity levels while sticking with the defendant; in Election Law, some judges might have likewise elevated their clarity assessments but switched to the plaintiffs' side.

D. Implications

1. Patterns and Limitations

Our survey indicated that successful lexical ordering would exact sacrifices. Many of the surveyed judges had general views about statutory vagueness and legislative history's usefulness such that a clarity test for access to lower-tier sources would generate unwelcome results.

Our experiment then showed mixed judicial success at lexical ordering. The Trade Name results are a success story, largely. Perhaps some judges failed to ignore the committee reports when required by the instructions but such failures are not statistically apparent. Moreover, the judges' general views about statutory vagueness and legislative history did not help predict

clarity responses in our models, which is consistent with judges concentrating on the instructions and the top-tier sources at hand. The Election Law results, in contrast, include stronger warnings of failure. The agency position seems to have influenced some judges' clarity assessments, and perhaps their final judgments, when only top-tier sources should have been considered.

Our experiment cannot fully explain these patterns. But, even at this stage, we can suggest possible predictors of failure. Other commentators have maintained that (1) *the case's* perceived importance and ideological charge can influence whether judges depart from plain meaning;[210] and (2) *the source's* credibility and emotional charge can influence whether judges consider the information.[211]

On case features, our results arguably are consistent with instructions weakening as case importance and ideological charge rise. The Election Law vignette was drafted to be more ideologically charged and less routine than Trade Name.[212] We cannot be sure from the data, but it would not be surprising if judges care more about a voting-rights dispute between employees and employer than a product-naming dispute between two unknown companies. Note that this is not a purely attitudinal account. The weakening of lexical ordering instructions would depend partly on exposure to an additional source of law.

On source features, our experiment can only foreground possibilities. First, agency positions might be more attractive than legislative history, at least in the form of smoking-gun snippets in committee reports that are not integrated with a deeper argument about the statute.[213] Judges might be less likely to ignore the decisiveness and possibly the expertise in agency positions—even if judges are committed to agency-deference doctrines as a matter of binding law.[214] If so, the lexical ordering element of *Chevron* doctrine would be especially vulnerable.[215] Second, judges might have been attracted

210 *See* Schauer, *supra* note 99, at 247–49 (suggesting that justices might use plain meaning readings as a strong factor in decisionmaking when the case is somehow not interesting or less politically charged); *cf.* Ryan D. Doerfler, *High-Stakes Interpretation*, 116 MICH. L. REV. 523, 530 (2018) (contending that courts tend to resist finding statutory clarity when the case "matters specially").

211 *See* Wistrich et al., *supra* note 152, at 1275–76.

212 *See supra* note 191 and accompanying text.

213 I thank William Eskridge for raising a question along these lines.

214 In Gluck and Posner's illuminating interviews with a mostly nonrandom sample of federal circuit judges, several respondents doubted that the Supreme Court had authority to bind them to broad interpretive rules. *See* Gluck & Posner, *supra* note 16, at 1306–07. But some of those decentralist judges made exceptions for certain canons such as lenity, and "every interviewed judge told us [the *Chevron* doctrine] is binding even if they disagreed with it." *Id.* at 1345 (emphasis omitted). It seems that about half of the respondents indicated that they consider legislative history (only?) if statutory text is unclear, *see id.* at 1316, but the reported results do not have a count of judges who support or accept such lexical ordering.

215 To the extent that *Chevron* imposed lexical ordering but did not change the likelihood of judicial deference, one explanation is that judges often cannot ignore the agency's position at Step One (or elsewhere). If so, converting *Chevron* to a supermajority voting

to the agency's proplaintiff position rather than the agency's status. Perhaps more judges were on the verge of adopting the plaintiff's position at baseline, and thus exposure to a proplaintiff lower-tier source became too attractive to ignore.[216] Third, other source content might matter. The agency position in Election Law included a short policy rationale while the committee reports in Trade Name did not. Perhaps the same Election Law message from a litigating party would have similar influence.[217] Our experiment was designed to implicate these potential influences without disentangling them.

With adequate numbers of judge participants, we can learn more through additional experimental variation. Varying the lower-tier source type, direction, and other content, as well as the importance and ideological charge of the cases, would help identify any influence of these factors. Our vignettes' order of presentation was fixed and, likewise, it could be varied to examine order effects. Furthermore, we imposed a modest cognitive load. The vignettes were short and judges were not asked to write opinions or deliberate in groups. This design is convenient and follows several previous studies,[218] plus most judges face resource constraints that make rapid judgments realistic.[219] Even so, the effects of docket load, case complexity, and time pressure are independently important topics.[220]

Beyond standard external validity concerns, our experiment relied on relatively less experienced judges.[221] Experience might matter. If, for instance, greater judicial experience increases a person's proficiency at lexical ordering, then we have another reason to confront the hard trade-offs in Part II. Moreover, our participants were appellate rather than trial judges, mostly state rather than federal judges, and, for the sake of anonymity and response rates, they were not asked demographic, policy, or political affiliation questions. We know little about the participants' general views on statutory interpretation. A different mix of judges and additional data on each judge could yield further lessons.

---

rule that asks judges for "the best interpretation of the statute," Jacob E. Gersen & Adrian Vermeule, Chevron *as a Voting Rule*, 116 YALE L.J. 676, 679 (2007), would likely not produce very different judicial outcomes compared to a pre-*Chevron* multifactor test.

216 Our control group was far more likely to be persuaded by the plaintiffs' position in Election Law than in Trade Name.

217 Also, to reflect contemporary doctrine, the instructions stated that judges "should consider" the committee reports if the statute was otherwise relatively unclear, but that judges "should defer" to the agency interpretation. *See infra* Appendices A & B. Perhaps these formulations differently influence judges' likelihood of ignoring lower-tier sources while working with top-tier sources—although that influence is not obvious theoretically.

218 *See supra* Section III.A (describing experiments by the Guthrie group, the Kahan group, and others).

219 *See* Aaron-Andrew P. Bruhl, *Hierarchy and Heterogeneity: How to Read a Statute in a Lower Court*, 97 CORNELL L. REV. 433, 470–72 (2012) (distinguishing most courts from the U.S. Supreme Court).

220 *Cf.* Spamann & Klöhn, *supra* note 170, at 268–73 (finding, in a labor-intensive decision setting, that judges were unable to ignore simple extralegal case facts in deciding an international criminal dispute).

221 *See supra* notes 187–88 and accompanying text.

2. Failure's Types and Success's Burdens

In any event, we can be sure that lexical ordering fails in more than one way, and that not every failure is equally important. Beyond effects on clarity assessments, which we saw in Election Law, our experiment focused on serious failures involving final judgments. Our tests did not confirm that judges conveniently used the clarity test to make attractive sources legally relevant and unattractive sources legally irrelevant. Those possibilities deserve ongoing investigation but they are not obvious in our results. On the other hand, some judges in Election Law might well have flipped judgments at relatively high clarity levels. This would be an acute form of failure in which a lower-tier source pushes some judges from near one end of the clarity-plus-judgment spectrum toward the other.

Tainted clarity assessments are less dramatic failures, but they can be troubling. Consider the possibility of judges building "clarity cushions,"[222] which avoid recognition of close calls that tax judges cognitively or in exposition. This could happen when judges find a lower-tier source controversial, discordant, or otherwise problematic. Think about the committee reports in Trade Name. Not only would some judges find this category of source suspect, but many would view the reports as conflicting with top-tier sources.[223] Perhaps some treatment judges stuck with the defendant and elevated their clarity assessments even higher, to make the proplaintiff committee reports more plainly irrelevant as a matter of law.

Clarity cushions might appear when final judgments flip, too. In Election Law, judges seemed inclined toward the agency position, and a high clarity assessment would ensure the legal irrelevance of that source. That seems inconvenient. But some judges who perceived a modest degree of clarity for the defendant in the top-tier sources might have been encouraged to take the plaintiff's side after seeing the agency's position, without relying on it explicitly.[224] As their judgments flipped, an elevated clarity assessment would position these judges to maintain that the statute already plainly supported the plaintiffs, and that the agency's position either need not be addressed,[225] or provides an alternative ground for judgment.[226] Certitude is a convention for some judges.[227]

---

222 *See supra* subsection III.C.3. ("Other Possibilities").

223 In Trade Name, confidence among the treatment judges seems *lower* than the control judges. *See supra* Table 2.

224 In Election Law, confidence among the treatment judges seems *higher* than the control judges. *See supra* Table 2.

225 *Cf.* Roberts v. Sea-Land Servs., Inc., 566 U.S. 93, 100, 113 n.12 (2012) (finding that a statute clearly favored the agency's position, and thus not addressing whether such position was eligible for *Chevron* deference).

226 *Cf.* Bimini Superfast Operations LLC v. Winkowski, 994 F. Supp. 2d 106, 125 (D.D.C. 2014) (indicating that the court would defer to the agency's position even if the statute were unclear).

227 *Cf.* Dan Simon, *A Third View of the Black Box: Cognitive Coherence in Legal Decision Making*, 71 U. CHI. L. REV. 511, 513 (2004) (discussing coherence-based reasoning that

Thus several types of failure can be linked to plausible theory, suggestive evidence, or both. Lower-tier sources might improperly affect clarity assessments, final judgments, or both. When judgments are improperly affected, the failure is serious and covert, and it might or might not equate with an aboveboard all-inclusive approach to sources.[228] When judgments are not affected, the failure is less serious but no less covert, and the lexical ordering effort might be wasteful to judges and probably will be misleading to observers. For example, when clarity assessments remain high yet affected by controversial lower-tier sources, judges avoid overtly grappling with those sources and likely will misstate the difficulty of the decision. Whether or not the parties mind, those opinions might be cited and used as models of interpretive reasoning going forward. The probabilities for each failure type remain to be seen.

None of this should obscure the signs of judicial success in our results. Many of our tests did not show judges manipulating clarity tests or shifting judgments in defiance of their instructions. At the same time, we should know that the hard trade-offs in Part II arise precisely when judges succeed at lexical ordering. "Success" means that decision quality and decision costs tend to fall in a mediocre range compared to prominent alternative methods, even as decisiveness is maximized. This compromise should not be accepted without careful reflection. While not every lexical ordering failure is important to the legal system, any success is reason to wonder whether lexical ordering is worthwhile.

3. Toward Reform

We can close by recognizing some motivations and options for reform today. Additional research is appropriate before reaching definitive conclusions, but waiting is not entirely possible for judges. The live methodological choices will not disappear as we build knowledge, and lexical ordering's nationwide spread increases the urgency of reconsideration.

At minimum, nobody should casually assume that a judiciary's lexical ordering efforts will succeed or fail. Proponents are not entitled to believe that courts have reached a compromise with predictable effects across the statutory docket, and critics are not entitled to disregard lexical ordering avowals as so much ineffectual formalism. Success will be mixed. And success might well be most likely in the heartland of lower court dockets, where most judges professionally process tall stacks of apparently ordinary cases.

---

yields confident conclusions and its limits); Brad Snyder, *The Judicial Genealogy (and Mythology) of John Roberts: Clerkships from Gray to Brandeis to Friendly to Roberts*, 71 OHIO ST. L.J. 1149, 1188 n.235 (2010) (noting a quip, attributed to Justice Brandeis, "that the difficulty with this place is that if you're only fifty-five percent convinced of a proposition, you have to act and vote as if you were one hundred percent convinced").

228 Although possible, we cannot yet demonstrate that failed lexical ordering equals an all-inclusive approach. Our experiment compared *Caminetti*-style lexical ordering efforts with flat exclusion of sources, not an *American Trucking*–style inclusive approach.

Those cases add up, along with the complications and downsides of lexical ordering efforts.

When it succeeds, lexical ordering imposes an odd mix of consequences: middling decision quality, middling decision costs, and superior judicial decisiveness. That combination lacks a straightforward justification on interpretive principle,[229] and it is obviously debatable in pragmatic terms.[230] Surely it supplies no firm ground for a nationwide hammerlock on interpretive method when sources of any substantial value are involved. Even complete judicial success at ignoring lower-tier sources when instructed leaves other hard jobs unfinished, including the development of an acceptable clarity test, the assignment of sources to tiers, and the weighting of sources within tiers. When flushed out, lexical ordering is a complex methodological option with controversial effects and no easy claim to superiority.

Concentrating on the value of public relations or compromise does little to improve the claim.[231] It is hardly obvious that existing "clarity" tests can build much confidence in the rule of law or anything else.[232] To the extent that observers do not understand the trade-offs, any resulting confidence probably is unearned. Success will be mixed, anyway, and so nobody should be uniformly impressed by the instruction. Indeed, the mixture of success and failure could specially threaten confidence levels. If instructions truly are weakest in cases that are ideologically charged or high stakes, then observers are most likely to witness or suspect methodological departures in the highest profile cases. And for judges seeking compromise and guidance, lexical ordering would then achieve the least influence in the cases that probably need it most.

If nothing else, courts and others probably should renew an attitude of respect for judges who want to organize their thinking differently from the orthodox compromise that lexical ordering represents at the moment. Whether proexclusion or proinclusion, simpler and more extreme methodological alternatives for statutory cases might justifiably gain ground, or just tolerance, once the trade-offs and implementation challenges for lexical ordering are better understood. In this regard, methodology would be a relatively decentralized judge-by-judge choice. Somewhat similarly, we could renew respect for today's outlier jurisdictions that do not attempt to position legislative history or agency positions in lexically inferior tiers. Past efforts to restate and spread a mainstream view on statutory interpretation have had limited influence at the state level.[233] Understanding the hard choices and uncertainties that surround lexical ordering provides reason to appreciate

229 *See supra* Section II.A.

230 *See supra* Section II.B.

231 *See supra* Section II.C.

232 *See supra* text accompanying note 89. In our survey and experiment, judges were not all over the map in their estimations of statutory clarity, but there was noticeable variation in Figures 1–3.

233 *See supra* note 59 (discussing the American Bar Association's uniform state laws efforts).

interjurisdictional diversity. Not every state should be asked to adopt plain meaning rules. The alternatives remain all too plausible.

For unshaken supporters of centralized guidance, a more ambitious response is to shore up the lexical ordering instructions by restructuring the judicial decision environment—which, as far as we know, has never happened. Probably unsurprisingly. To ensure that lower-tier sources do not improperly influence judges, other staff could be assigned the task of withholding that information unless and until judges conclude that the statute is otherwise unclear. True ignorance of a source guarantees its powerlessness. But even setting aside the special problem of keeping judges in the dark about agency positions, reengineering the decision process would require more than extra time and faithful staff within chambers. For appellate courts, it would require coordination across chambers so that no judge on a shared panel is prematurely exposed to lower-tier sources by the other judges. Not only the briefing and bench memorandum process would need restructuring, but also the mechanics of oral argument, deliberation, and opinion writing. Without such architectural safeguards, judges might be left to train and concentrate as well as they can, with limited prospects for success. The truly effective fixes are costly and perhaps unrealistic today. For those who already are unsettled by lexical ordering's trade-offs, the prospect of additional effort to implement the instructions should turn their attention to other options.[234]

A more resigned yet thoughtful response is to abandon lexical ordering efforts in statutory interpretation—which certainly has happened before. As many New Deal and post–New Deal judges became interested in additional information for deciding statutory cases and dissatisfied with plain meaning barriers thereto, they cast aside what they thought was an artificial hindrance to thoughtful judgments. They confirmed legislative history and agency positions as potentially important factors in their decisions, among other considerations. And as some Reagan-era and post-Reagan judges became wary of the amount and quality of sources such as floor statements and committee reports, they cast aside what they thought was distracting and inconclusive information for understanding statutes. Both positions are extreme in a sense, but they also share the simplifying feature of inquiry freed from a clarity test, however elaborated. Neither is a good candidate for judiciary-wide compromise in most jurisdictions at this date, and both have their own draw-

234 This is not to conclude that convenient case-closing tiebreakers are unjustifiable. It is to suggest that attempting to assign lexical inferiority to a source category with any significant net value is problematic. However, our findings do begin to raise questions about part of the theory for choosing tiebreakers. Demoting agency interpretations is theoretically good for judicial decisiveness, generally speaking, while demoting legislative history is theoretically good for lowering decision costs. *See supra* text accompanying note 135. If agency positions often cannot be ignored, then their tiebreaking function is threatened, assuming no double counting. *See supra* note 136. But again, our experiment cannot confirm that different source categories have different effects on judges' ability to lexically order successfully. That question is open for future research.

backs measured in decisiveness and more. But those options do not labor under the complications imposed by the curious logic of lexical ordering.[235]

CONCLUSION

Today the defense of lexically ordered statutory interpretation is not much better than familiarity plus optimism. Its logical structure is popular in form but sometimes fragile in practice, and challenging to justify in any event. Under certain conditions, interpreters surely can carry out these instructions; under other conditions, the instructions will fail or make no difference. When the instructions are followed, the trade-offs often are troubling. Judges may avoid those particular trade-offs either by truncating the analysis, if the information can be flatly excluded, or instead by integrating all relevant information into the mix. Although we still have much to learn about lexically ordered interpretation, its current popularity far exceeds our ability to confidently justify its continuation. Indeed, the spread of lexical ordering in statutory cases should generate powerful demands for more knowledge about its real-world operation, along with respect for judges and judiciaries who would depart from today's orthodox ordering—the core problems of which are, by now, almost clear.

235 Judges who want to flatly exclude sources need a decision architecture that will help them ignore sources. If some appellate judges on the same court hold to more inclusive approaches, then proexclusion judges face a design problem partly shared with lexical ordering proponents. Proinclusion judges face a different problem: ensuring that all sources are considered thoughtfully, especially when the workload feels heavy.

Appendix A

Case #1: Trade Name

You are a judge assigned to decide the following case:

Company X is developing a drug that it calls Nova, and it has shipped the drug with the Nova label to several laboratories to test the drug's safety. Company Y recently began selling a different drug that it, too, calls Nova.

A statute prohibits, as an unfair trade practice, the use of "a product name that already has been used in commerce openly by another person or company."

Company X asserts that its drug has been "used in commerce" according to the ordinary meaning of those words. The company quotes dictionaries to show that the word "commerce" may be used to mean "activities that relate to the buying and selling of goods and services." The company emphasizes that its drug already has been shipped for safety tests, although the company admits that the drug has not been bought or sold.

Company Y asserts that Company X's drug has not yet been "used in commerce." Company Y points out that the dictionary definition quoted above is a secondary definition. The first dictionary definition of "commerce" is "the buying and selling of commodities." The company contends that the word "openly" in the statute provides context for concluding that this statute follows the primary definition of "commerce," and does not protect products shipped for safety testing before hitting the market. In other sections of the same statute regulating unfair trade practices, "used in commerce" has been interpreted by courts to mean the actual buying and selling of commodities — and not activity merely related thereto.

[CONTROL: How clear is the statute as applied to the facts of this case, on a scale of 1 to 5?]

[TREATMENT: Company X replies by stressing the committee reports produced in the legislature. Those reports describe "used in commerce openly" as "intended to include commercial sales along with less traditional product uses such as distribution for test marketing and safety testing." No other legislative history contradicts these reports. The courts in your jurisdiction have decided that judges should consider committee reports if but only if the other sources of statutory interpretation, taken together, leave the statute's meaning fairly or totally unclear as applied to the facts of the case.

How clear is the statute as applied to the facts of this case, without considering the committee reports, on a scale of 1 to 5?]

1 = Totally unclear
2 = Fairly unclear
3 = Somewhat clear
4 = Fairly clear
5 = Totally clear

For whom would you enter judgment: Company X or Company Y? . . . .

APPENDIX B

Case #2: Election Law

You are a judge assigned to decide the following case:

Twenty employees earn hourly wages at a manufacturing plant. They took time off from work to vote in the last election. Their employer did not pay them for the hours that they were away from work.

A statute declares that "employers shall allow each employee the time needed to vote during the workday, without penalizing them with a reduction in wages."

The employees assert that they suffered a "reduction in wages" according to the ordinary meaning of those words. The employees quote dictionaries indicating that "reduction" may be used to mean "making a specified thing smaller or less." The employees claim that their wages were made less by their employer because they left work to vote.

The employer agrees that the quoted dictionary definition is an ordinary meaning of "reduction," but asserts that the "specified thing" that must not be reduced is wages paid for work actually done. The employer contends that the word "penalizing" in the statute helps show the context in which "reduction" was used, and that not being paid while absent from work is not really a penalty. The employer also points to statutes addressing jury duty and medical appointments. In those statutes, prohibitions on "a reduction in wages" have been interpreted by courts to mean a reduction in wages paid for work actually done.

[CONTROL: How clear is the statute as applied to the facts of this case, on a scale of 1 to 5?]

[TREATMENT: The employees reply by emphasizing the position of the agency charged with administering the relevant statute. The agency has concluded that the statute "is intended to allow employees the freedom to go to the polls without losing any money that they could have earned at work, just as business owners who lack day-to-day job duties are usually free to go to the polls without financial loss." The courts in your jurisdiction have decided that judges should defer to an agency's position on the meaning of a statute if but only if the other sources of statutory interpretation, taken together, leave the statute's meaning fairly or totally unclear as applied to the facts of the case.

How clear is the statute as applied to the facts of this case, without considering the agency's position, on a scale of 1 to 5?]

1 = Totally unclear
2 = Fairly unclear
3 = Somewhat clear
4 = Fairly clear
5 = Totally clear

For whom would you enter judgment: the employees or the employer? . . . .