In statutory cases, judges often indicate that interpretive sources should be lexically ordered. That is, a limited set of top-tier sources, if adequately clear, are supposed to establish a statute’s meaning regardless of any other consideration. Familiar examples include plain meaning rules that demote legislative history and agency interpretations into tiebreaking positions. This Article scrutinizes lexical ordering’s nationwide spread, its trade-offs when followed, and the problems for its implementation. As for trade-offs, the analysis spotlights not only decision quality and decision costs but also judicial decisiveness. As for implementation, the Article reports results from a survey and an experiment conducted with approximately one-hundred appellate judges. These judges showed mixed success in lexically ordering sources, and some of the apparent effects of exposure to lower-tier sources were not easily predicted. In the final analysis, judges might be able to enhance the simplicity and quality of decision making in statutory cases by discarding lexical ordering’s complex, problematic, and sometimes ineffectual logical architecture. But there is more to learn.
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

Sidelined at the Supreme Court for decades, lexical ordering has reentered the core conventions of statutory interpretation in the United States. That is, most judges are supposed to decide whether a statute is clear using a limited set of top-tier interpretive sources and, if so, apply this meaning; if the statute remains unclear, lower-tier sources may or must be considered. Lower-tier sources are neither excluded nor included outright, but rather held aside just in case. Examples familiar to lawyers involve plain meaning rules that categorically demote—without killing off—considerations such as legislative history, deference to administrative agencies, and the rule of lenity.1 Some version of this analytical strut is now planted in the approved interpretive method of nearly every court system in the country. A popular judicial turn of phrase is that legislative history “is meant to clear up ambiguity, not create it.”2

Actually, the idea of lexically ordered interpretation is no younger than Blackstone’s Commentaries. In 1765, Blackstone listed secondary considerations in statutory interpretation that were supposed to matter only if “words happen to be still dubious.”3 His lower-tier sources included statutes on the same subject, along with the reason and spirit of the statute at issue.4 Lexically ordered statutory interpretation, which can take many forms, was endorsed at the Supreme Court by 1920.5 But that approach always had competitors. Sutherland’s Statutes countered with an apparently wide-open aggregation of legitimate interpretive sources. The 1943 edition of what had been J.G. Sutherland’s treatise pronounced that “statutory interpretation . . . is a fact issue. Where available, the courts should never exclude relevant evidence on that issue of fact.”6 In this regard the treatise was less Blackstone and more Holmes,7 as were many judicial opinions starting in the 1940s and lasting into the 1970s.8

1 See infra Part I.B.
3 1 WILLIAM BLACKSTONE, COMMENTARIES ON THE LAWS OF ENGLAND *59–*61 (1765) (Stanley N. Katz ed., 1979).
4 See id.
5 See, e.g., Caminetti v. United States, 242 U.S. 470, 490 (1917); cf. United States v. Fisher, 6 U.S. 358, 399 (1805) (Marshall, C.J.) (“Where a law is plain and unambiguous, . . . the legislature should be intended to mean what they have plainly expressed, and consequently no room is left for construction.”); id. at 386 (stating that “[w]here the intent is plain, nothing is left to construction” and then, in apparent contrast, that “[w]here the mind labours to discover the design of the legislature, it seizes every thing from which aid can be derived”).
6 2 J.G. SUTHERLAND & FRANK E. HORACK, JR., STATUTES AND STATUTORY CONSTRUCTION § 4502, at 317 (1943 ed.).
7 See Boston 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 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.”).
Of course Blackstone did not evaluate the trade-offs and implementation challenges for lexically ordered interpretation. Nor have the rest of us done much better since then, even as the idea returned to prominence by the 1990s. Perhaps we have become all-too familiar with lexical ordering as a methodological compromise. One side’s persistent support for using certain interpretive 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 on decisions. Instead, all interpreters are asked to demote the controversial sources into a lower tier in the decision tree—with the logical hinge to the demoted sources turning on each interpreter’s sense of “clarity,” which is to be developed while ignoring the demoted sources.

This compromise is complex, sometimes ineffectual, and problematic regardless. Lexical ordering taxes any judge who tries hard to follow the instructions, including judges who prefer to combine all sources into one mix and judges who prefer to flatly exclude lower-tier sources. If implemented successfully, any upside to lexical ordering might not be worth the effort in view of the downside risk that useful information in lower tiers will be ignored.

On the other hand, lexical ordering instructions might not matter much in practice. Last year, Judge Kavanaugh reasserted that there is nothing clear about clarity tests for statutes. Other critiques of plain meaning rules include Murphy, supra note 8, at 1316 (“One can rationalize adoption of the British rule on legislative history [excluding it], but not a selective rule of preclusion applicable only where the language is ‘unambiguous.’” (footnote omitted)).

9 Cf. Adrian Vermeule, Interpretive Choice, 75 N.Y.U. L. REV. 74, 82 (2000) (indicating that lexical ordering is an unavoidable methodological question for statutory interpretation that deserves serious consideration in light of high theory and limited evidence). The beginnings of the analysis on trade-offs presented here 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]. That article did not confront implementation challenges or the relationship between those challenges and trade-offs. For a recent contribution to the normative evaluation of lexically ordered interpretation, with several incisive observations that are broadly consistent with Tiebreakers, see William Baude & Ryan Doerfler, The (Not So) Plain Meaning Rule, 84 U. CHI. L. REV. 539 (2017) (discussed below). Older critiques of plain meaning rules include Murphy, supra note 8, at 1316 (“One can rationalize adoption of the British rule on legislative history [excluding it], but not a selective rule of preclusion applicable only where the language is ‘unambiguous.’”) (footnote omitted).

10 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 version of textualism” that is prevalent in state law and that prioritizes certain forms of textual analysis over legislative history, and legislative history over certain judicial presumptions). Gluck’s case studies are real contributions, but that article does not tackle how or whether lexical ordering can be implemented effectively, nor the trade-offs at stake. The article’s descriptive account does refer to values such as stability and predictability, but makes no effort “to prove that methodology dictates outcomes.” Id. at 1770. The stability at issue in Gluck’s article seems to be expositional: how judges explain themselves, not necessarily how judges decide cases.

11 See infra Part II.B.1–2.

Thus judges might adopt convenient understandings of clarity that usually allow each judge to consider lower-tier sources when they want to and not when they don’t. Perhaps judges who tend to find statutory text clear also tend not to find legislative history useful, and vice versa. If so, the on-the-ground compromise in practice will drift away from lexical ordering and toward decentralized judge-by-judge choices of method. And perhaps the inability of the clarity test to bite hard is what makes lexical ordering a popular compromise for statutory cases today.

Or judges might not be able to follow lexical ordering instructions when they should bite. “When the reading is done and the case has been analyzed and argued,” Judge Randolph asked more than twenty years ago, “how can it be said that the judge turned to the legislative history only after finding the statutory language ambiguous?”\textsuperscript{13} The same question can be asked about any lower-tier source—especially because those who favor lexical ordering have not offered realistic advice for ensuring that demoted sources have no influence on initial assessments of clarity, not to mention that judges who prefer to ignore lower-tier sources will consider them when the statute is otherwise unclear. True, implementation difficulties might soften the bite of the compromise, as supposedly demoted sources creep into higher tiers. But everything depends on exactly how the lexical ordering effort is designed and executed.

This Article attempts to make theoretical and empirical progress on all of the above ideas. Part I reviews the logic of lexical ordering and shows how far it has spread into the conventions of statutory interpretation in state and federal courts. The discussion focuses on the position of legislative history and agency interpretations, noting changes over time, outlier practices, and the leadership role of judges rather than legislatures. Part II identifies key trade-offs that judges should associate with lexically ordered interpretation, assuming the instructions are followed. Compared to aggregating all relevant source inferences, lexically ordering those sources risks reducing the quality of interpretation but could moderate decision costs and probably will increase decisiveness. Compared to excluding lower-tier sources, lexical ordering probably allows for higher quality interpretation and increased decisiveness, but also higher decision costs. These are generalizations, however, which depend on successful implementation in judicial decision-making, not only opinion-writing.

\textsuperscript{13} A. Raymond Randolph, \textit{Dictionaries, Plain Meaning, and Context in Statutory Interpretation}, 17 Harv. J.L. & Pub. Pol’y 71, 76 (1994); accord ESKRIDGE ET AL., supra note 9, at 438 (“[Y]ou can never know whether a plain meaning has been found without any reference to legislative history (for the judges have read that part of the brief).

Part III drops the assumption that lexical ordering instructions will be followed and turns to implementation problems. The discussion presents the results of a new survey and vignette experiment with sitting appellate judges. Certainly not every judge conveniently applies clarity tests to block or access lower-tier sources at will. The surveyed judges’ reported tendencies to find statutory text unclear are not significantly correlated with their reported tendencies to find legislative history useful. However, the vignette experiment suggests that judges only sometimes ignore the content of lower-tier sources while they work with higher-tier sources. Some judges seem attracted to—and others might be repelled by—lower-tier sources in ways that lexical ordering is supposed to prevent.

There is more to learn. For now, we can observe that lexical ordering instructions actually are likely to bite in some statutory cases, ruling out potentially useful information and posing trade-offs that are underappreciated—while in other cases the instructions will fail to guide decisions, becoming an unwelcome distraction if not worse. As we build theory and evidence to better identify these domains, we should doubt that lexical ordering’s current popularity in courts is justified based on current knowledge. In any event, lexically ordered interpretation is a subject that deserves more attention given its prominence and problems. We should be asking whether this compromise can be implemented successfully and, when it can be, whether the trade-offs are worth it.

I. CONCEPTS AND TRENDS

A. Lexical Ordering

Lexical ordering creates a particular kind of priority and a particular kind of condition for the consideration of various sets of information that might be used to make a single decision. It is a decisive logical architecture that dictates conditions on which information should be used, not simply a convenient sequence of analysis that never discards information. Higher-tier (lexically superior) considerations must be the exclusive basis for a given decision and trump any lower-tier (lexically inferior) considerations, which are excluded from the analysis unless the top-tier considerations are not adequate on their own to deliver a unique result. Part of lexical ordering’s burden is the task of determining, justifiably, what counts as an adequate basis for decision using only higher-tier sources.

14 See infra Part III.B & fig. 1.
15 See infra Part III.C, tbl. 1 & figs. 2–5.
16 The challenges for lexically ordered interpretation generally do not depend on which sources are placed in which tiers. True, each type of source—judicial precedent, canons, legislative history, agency positions, or some other source category—can present special challenges for judicial use. But this Article concerns the architecture of lexical ordering, and is not an attack on or a defense of any particular type of interpretive source. Extensions of the analysis to other fields of law—contracts, patents, treaties, and elsewhere—are possible.
17 Sequencing of sources is necessary to interpretation; lexical ordering is not. Lexical ordering is designed to foreclose the consideration of a category of information, at least sometimes, while conscientiously sequencing information implies no such thing. 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 potential cognitive effects of sequencing choices, including conditions in which a “text first” sequencing rule can backfire).
18 See Samaha, Tiebreakers, supra note 9, at 1669 (strictly defining a tiebreaker as a lexically inferior decision rule).
This might look a bit complicated, and it is, but the logic is no more foreign than the rules for alphabetizing. 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. Nobody doubts the usefulness of lexical ordering for certain purposes, such as organizing book shelves where books are still in use. Whether lexical ordering is good for much else, however, ought to be controversial.

In the field of statutory interpretation, we face long-running and unavoidable debates about the legitimacy and value of interpretive sources. Some people contend that legislative history such as committee reports and floor statements should never influence judicial decisions in statutory cases, while others maintain that such legislative history should be one consideration within a larger mix and for whatever it’s worth. Parallel debates take place over the appropriate influence of an administrative agency’s position on a judge’s interpretation of a statute. These debates are necessary, even healthy.

But none of these positions require lexical ordering. As a matter of principle, interpreters can exclude illegitimate, irrelevant, or otherwise disfavored sources of information, and they can include various sources to build an aggregate mix of inferences while adjusting the weight assigned to each source—all without lexical ordering. Lexical ordering neither fully excludes nor fully includes lower-tier sources at the outset. With steely determination, lexical ordering aims to prevent the aggregation of implications across otherwise relevant sources, but then, with sudden inclusiveness, tries to reintegrate those sources based on the operation of a specified logical hinge. Potentially useful information is supposed to be channeled away from the interpreter, sometimes, while, at other times, that same information floods back into the analysis.

Despite the complications, lexical ordering has taken hold of many judges and it can be understood as a compromise. Not a compromise involving reduced weight for controversial

19 See JOHN RAWLS, A THEORY OF JUSTICE 42–43 & n. 23 (1971) (analogizing to alphabetizing in describing the categorical normative superiority of certain considerations in evaluating political systems). On the early development of actual alphabetical arrangements, see LLOYD W. DALY, CONTRIBUTIONS TO A HISTORY OF ALPHABETIZATIONS IN ANTIQUITY AND THE MIDDLE AGES ch. II (1967).
22 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 Kevin M. Stack, An Administrative Jurisprudence: The Rule of Law in the Administrative State, 115 Colum. L. Rev. 1985, 1998 (2015) (observing that many people view the question whether a statute is clear enough to foreclose deference as one that “a court could only ask de novo,” though exploring the alternative view), with Peter L. Strauss, “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), and Jack M. Beerman, End the Failed Chevron Experiment Now, 42 Conn. L. Rev. 779, 779 (2010) (advocating a revival of pre-Chevron practice of multifactor inquiry).
23 When they do, issues of how to weight those sources will return as well.
24 Which is not to assert that lexical ordering always lacks a principled defense apart from the benefits of compromise. See infra text accompanying notes 98–100 (referencing John Manning’s defense under a particular form of purposivism).
sources within an aggregated mix of considerations (although the ex ante expected influence of such sources will decline upon successful implementation), nor a compromise involving a strong presumption in favor of relying on uncontroversial sources alone. Instead the compromise involves judges developing higher and lower tiers of sources, and setting “clarity” as the logical hinge between them. Understanding exactly how judges operationalize these clarity hinges is crucial to evaluating the compromises, which implicate difficult trade-offs when implemented successfully. Neither the hinge nor the trade-offs are anything like organizing a bookshelf alphabetically. The stakes can be high and the trade-offs unwelcome when lexical ordering is poorly designed or deployed, a point to which we will return shortly.

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 early twenty-first century, lexically ordered statutory interpretation became trendy. The discussion here concentrates on two prominent forms of lexical ordering—involving legislative history and agency interpretations—but the idea is more widespread than even this discussion will indicate.25 Also worth underscoring is that the most recent rise of lexical ordering in statutory cases has been driven by judges, not legislatures. The state-level construction acts that address lexical ordering are few, and a few of those statutes have been basically ignored by judges.

1. Legislative history

Congress has not given federal judges an overarching construction act to consult, but federal judges have largely embraced lexical ordering in federal statutory cases. An example involves legislative history in the forms of committee reports, committee testimony, and floor statements. Federal judges often tell us that the implications of legislative history are irrelevant unless the statute’s meaning as applied to the case at bar is in some sense unclear. Most recently, Matal v. Tam26 observed that the Supreme Court could ignore arguments from legislative history because “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.’”27 True, Matal went on to conclude that the party pressing legislative history in the trademark statute had not come up with anything very telling.28 And some of the Court’s modern remarks about the role of legislative history—that these sources “need not” be considered when the statute is otherwise clear—are less than definitive endorsements of lexical ordering.29 But the Court’s overall message commits to the lexical inferiority of legislative history.

25 For the demotion of general statutory purpose and policy-related consequences into a lexically inferior position compared to a clear statutory meaning in federal courts, see Sebelius v. Cloer, 133 S. Ct. 1886, 1895–96 (2013); John F. Manning, The New Purposivism, 2011 SUP. CT. REV. 113, 126–29 (2011). For the deeper demotion of the rule of lenity to a last-ditch tiebreaker, see Ocasio v. United States, 136 S. Ct. 1423, 1434 n.8 (2016) (indicating that lenity kicks in only “after seizing everything from which aid can be derived”) (internal quotation marks and citation omitted).


28 See Matal, 136 S.Ct. at 1756.

29 E.g., 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 (2013) (similar); see also Murphy, supra note 8, at 1303–04 (suggesting that such
The history of this commitment is both long and unstable, though. A lexical ordering commitment appeared in Supreme Court opinions by the early twentieth century. In *Caminetti v. United States*, dealing with the interstate transportation of women for immoral purposes, the Court thought it “elementary that the meaning of a statute must, in the first instance, be sought in the language in which the act is framed.” This statement might look like a mere sequencing rule but the Court followed up with an endorsement of lexical priority: “if that [meaning] is plain, and if the law is within the constitutional authority of the lawmaking body which passed it, 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.” *Caminetti* can be contrasted with the *Holy Trinity* approach to statutory interpretation in the late 1800s, because the latter emphasized the law’s spirit over the law’s text while the former did the opposite. But on lexical ordering, both approaches are essentially in lockstep. The approaches differ over which considerations judges should place in which tier, not whether judges should construct tiers at all.

Lexical ordering was apparently replaced by the ultra-inclusive “no rule of law” approach of *American Trucking* in 1940. Emphasizing legislative purpose, the Court’s opinion declared 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.’” This add-it-all up approach to interpretive sources was confirmed by a unanimous Supreme Court as late as 1976, when a lower court was rebuked for saying exactly what the high Court would later come to endorse again: that judges “need not resort to legislative history” when a statute is “plain and unambiguous.” Judges were not supposed to exclude the possible insights gained from legislative history based on supposedly plain meaning.

But by the early 1990s, lexical ordering for legislative history reappeared in Supreme Court opinions, as John Manning has observed. This development should be understood

“need not” language is not a plain endorsement of a plain meaning rule).

30 242 U.S. 470 (1917).
31 Id. at 485.
32 Id.
33 See *Holy Trinity Church 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.”).
37 See *Train*, 426 U.S. at 10 (quoting *American Trucking* and citing Murphy’s article); id. at 11–24 (rellying heavily on a committee report and floor debates to limit otherwise overlapping agency jurisdiction).
38 See Manning, *supra* note 25, at 126 (finding that an approach that promoted “the unyielding quality of a semantically clear statutory text . . . began to take shape the last year of the Burger Court”); Vermeule, *supra* note 34, at 183 (presenting a similar timeline toward literalism); see also West Virginia Univ. Hospitals, Inc. v.
against concentrated attacks on the intelligibility, reliability, and legitimacy of legislative history as a reflection of collective congressional intent. Those attacks certainly influenced judicial thinking, but not enough to flatly eliminate legislative history in all cases. Yet the demotion of legislative history into a lexically inferior tier of sources 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. In 2011, Justice Kagan added, “Legislative history, for those who take it into account, is meant to clear up ambiguity, not create it.”

State courts are where most judicial interpretation takes place in this country. While their officially approved methods vary, the vast majority of state court judges tell us that they push legislative history into a lower tier of considerations. Thus the Michigan courts have said that, when statutory text is found unambiguous, “the examination of legislative history ‘of any form’ is not proper.” Sometimes the state court’s position is not so categorical, as when California’s high court recently declared a statute clear and that “we need go no further”—then proceeded to check up on relevant legislative history. Nonetheless, most state courts align with the federal courts on this issue.

Only a few state courts are now bold enough to overtly reject the placement of legislative history in a lexically inferior position. New York courts remain outside the contemporary trend, for instance, although the State’s position has not been fixed for extended periods. Recently the high court reaffirmed that, “[a]lthough the plain language of the statute provides the best

---

41 The observations made in this discussion are based on a review of state construction acts and state supreme court opinions in the past five years or more. A spreadsheet of authorities is available from the author. An important start on such state-level research is Gluck, supra note 10, at 1844 & n. 353 (collecting cases and offering a tentative generalization about state court practices, alongside careful case studies).
43 Scher v. Burke, 395 P.3d 680, 688 (Cal. 2017) (proceeding to rely on a legislative counsel's digest, the enrolled bill memorandum to the governor, and committee analysis). Also consider the odd position that legislative history cannot overcome plain statutory text, and yet can be used to confirm a meaning that is already supposed to be plain. See Parker Land & Cattle Co. v. Wyoming Game & Fish Com'n, 845 P.2d 1040, 1044 (Wyo. 1993) (acknowledging occasional examples of this practice), quoted in State ex rel. Alden v. Kirchhefer, 357 P.3d 1118, 1127 n.1 (Wyo. 2015).
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.’44 Alaska’s judiciary is a resolute outlier on this matter, having articulated a “sliding scale approach” to plain meaning in conjunction with other sources,45 while New Mexico’s high court acknowledges extraordinary situations in which plain meaning “must yield” to indications from legislative history.46 But in the past few years, only a couple other state supreme courts have expressed even limited or inconsistent support for legislative history entering the top tier of sources.47

Occasionally the stakes for lexical ordering are low, though. “Legislative history” can be a very limited set of materials at the state level even for modern enactments. In terms of official state records, legislative history might refer to earlier enacted versions of the state code and really nothing else. This is partly because a few states do not maintain official legislative history in the sense of witness testimony, legislator debates, committee reports, or legislative staff summaries.48 In these jurisdictions, there is not much legislative history to look at even if one would like to see it. But note that lexical ordering is the prevailing approach in most states regardless. That is, some states demote earlier enacted versions of their statutes into a lower tier of interpretive sources, having no other kind of official legislative history to demote.49

Regardless, judges rather than legislatures lead the way. Most state codes of construction do not plainly speak to the question whether legislative history should be lexically inferior to other sources. Legislative silence on this point is not entirely for lack of trying. For its part, the American Bar Association’s National Conference of Commissioners on Uniform State Laws

---

44 Kimmel v. State, 29 N.Y.3d 386, 397 (2017) (citation omitted) (defending this inclusive approach serving the court’s goal of implementing the state legislature’s intent). Compare the less committed position on inclusion in Avella v. City of New York, 29 N.Y.3d 425, 438 (2017) (indicating that the court “need not” consider legislative history because of the statute’s plain language, but proceeding to consider it anyway), and contrast the earlier declaration in Washington Post Co. v. New York State Ins. Dep’t, 463 N.E.2d 605, 606 (N.Y. 1984) (“When the plain language of the statute is precise and unambiguous, it is determinative.”).

45 See Alaska Miners Ass’n v. Holman, 2017 WL 2609533, at *4 (Alaska 2017) (indicating that the court considers legislative history along with statutory text and purpose); Muller v. BP Exploration (Alaska), 923 P.2d 783, 788–89 (Alaska 1996) (rejecting a simple plain meaning rule in favor of “a sliding scale approach”).

46 Fowler v. Vista Care, 329 P.3d 630, 633–34 (N.M. 2014) (“The plain meaning rule ‘must yield on occasion to an intention otherwise discerned in terms of equity, legislative history, or other sources.’” (citation omitted)).

47 For some tension in recent pronouncements within the same state, compare AIDS Support Group of Cape Cod, Inc. v. Town of Barnstable, 76 N.E.3d 969, 974–75 (Mass. 2017) (stating that legislative history is “not ordinarily” a proper source of construction if statutory language is unambiguous, and consulting such sources), and A.J. v. Eighth Judicial Dist. Ct., 394 P.3d 1209, 1213 (Nev. 2017) (three-judge panel) (stating that “ambiguity is not always a prerequisite to using extrinsic aids”) (internal quotation marks omitted), with People for the Ethical Treatment of Animals, Inc. v. Dep’t of Agr’l Resources, 76 N.E.3d 227, 234-35 (Mass. 2017) (“[I]f [statutory] language is ‘clear and unambiguous, it is conclusive as to the intent of the Legislature.’”) (citation omitted), and Public Employees’ Retirement Sys. v. Gitter, 393 P.3d 673, 679 (Nev. 2017) (“Only when a statute is ambiguous will this court ‘resolve that ambiguity by looking to the statute’s legislative history and construing the statute in a manner that conforms to reason and public policy.’”) (citation omitted).

48 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, 154 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) (“[I]t is possible that legislative history in West Virginia is minimal, at best . . . .”).

49 See, e.g., Ex parte B.C., 178 So.3d 853, 855–56 (Ala. 2015) (indicating that, if statutory text is ambiguous, courts may look to legislative history and referring to “the law as it existed prior to such statute’s enactment”).
recommended in 1993 that legislative history be demoted to a lower tier of sources.\textsuperscript{50}  Like the Conference’s 1965 effort at a uniform statutory construction act, which was not as clearly committed to lexical ordering, few state legislatures followed the 1993 proposal.\textsuperscript{51}

The construction acts that do speak to the proper role for legislative history are a bit mixed. The Texas code provision, which was adopted in 1985, might read as a demand that legislative history be included in a top-tier mix: “In construing a statute, whether or not the statute is considered ambiguous on its face, a court may consider among other matters . . . legislative history.”\textsuperscript{52}  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: “A court shall give the weight to the legislative history that the court considers to be appropriate.”\textsuperscript{53}  In contrast, the Pennsylvania code, which was adopted in 1972, obviously embraces lexical ordering for legislative history: It 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 pretext of pursuing its spirit,”\textsuperscript{54} and goes on to approve consideration of legislative history and other listed sources “[w]hen the words of the statute are not explicit.”\textsuperscript{55}  Several other state codes are equally or almost as explicit about demoting legislative history.\textsuperscript{56}

Whether these statutes matter is open to question. An on-point construction act has not always influenced even official judicial policy on legislative history,\textsuperscript{57} and we have yet to see evidence of an effect on case results. Not ironically, construction acts are statutes that must be “interpreted” by judges. Despite the permissiveness shown to legislative history by the Texas

\textsuperscript{50}  See Uniform Statute and Rule Construction Act §§ 18–20, 14 U.L.A. 406 (1993) (adopting lexical ordering for, among other factors, legislative history); id. § 20 cmt. (p. 36) (confirming that the “step-by-step” approach should be that “[t]o use the materials described in subsection (c) [such as legislative history] the construer must conclude that the steps taken under Sections 18, 19, and 20(a) and (b) still leave the meaning uncertain”).

\textsuperscript{51}  See Uniform Statutory Construction Act §§ 13 & 15, 14 U.L.A. 765–66 (1965) (offering presumptions and a nonexhaustive list of sources that may be considered when a statute is ambiguous, without expressly endorsing a plain meaning rule).

\textsuperscript{52}  TEX. GOV’T CODE § 311.023 (as amended by 1985 Tex. Acts, 69th Leg., ch. 479, § 1).

\textsuperscript{53}  ORE. REV. STAT. § 174.020(b)(3) (as amended by 2001 Ore. Laws ch. 438, § 1); see also id. (“[A] party may offer the legislative history of the statute. . . . A court may limit its consideration of legislative history to the information that the parties provide to the court.”). Note the labor-saving device offered to judges, who were granted permission to limit review to sources put forward by litigants.

\textsuperscript{54}  1 PA. STAT. & CONS. STAT. ANN. § 1921(b) (enacted by 1972 Pa. Pub. L. 1339, no. 290, § 3).

\textsuperscript{55}  Id. § 1921(c)(7).

\textsuperscript{56}  Three state codes have language akin to the two Pennsylvania clauses quoted above. See LA. REV. STAT. § 1:4 (first clause on letter and spirit); id. § 24:177 (second clause on legislative history for ambiguity); MINN. STAT. ANN. § 645.16 & 645.16(7) (both clauses); N.D. CENT. CODE § 1–02–05 (first clause); id. § 1–02–39.3 (second clause). Two other state codes seem substantively similar. See CONN. GEN. STAT. ANN. § 1-22 (enacted in 2003) (“If, after examining [statutory] text and considering such relationship [to other statutes], the meaning of such text is plain and unambiguous and does not yield absurd or unworkable results, extratextual evidence of the meaning of the statute shall not be considered.”); N.M. STAT. ANN. § 12-2A-20.C(2) (reserving “the purpose of a statute . . . as determined from the legislative or administrative history of the statute” for cases of residual uncertainty). Colorado has the first clause on letter and spirit without the second clause on legislative history. See COLO. REV. STAT. ANN. § 2-4-203(1)(c). Iowa and Ohio have the second clause without the first. See IOWA CODE § 4.6.3 (enacted in 1971); OHIO REV. COD. ANN. § 1.49(C).

\textsuperscript{57}  See Gluck, supra note 10, at 1771–91 (examining carefully state court opinions in five jurisdictions including Oregon and Texas, and showing apparent departures from state statutory instructions for interpretation).
code, which indicates that ambiguity is not a prerequisite to its consideration, the Texas Supreme Court nonetheless has declared that “legislative history cannot override a statute’s plain words.”58 The court recently suggested that the matter is beyond the legislature’s authority.59 Meanwhile, the Oregon Supreme Court saw its Legislature’s invitation to assign weight to legislative history and responded with a rule: “no weight can be given to legislative history” when statutory text can have only one meaning.60 Headstrong judicial leadership in statutory cases can run the other way, too. 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.61 Yet New Mexico’s Supreme Court has affirmed that “[t]he plain meaning rule ‘must yield on occasion to an intention otherwise discerned in terms of equity, legislative history, or other sources.’”62

2. Agency interpretations

With these trends in mind, agency interpretations can be discussed more briefly. For federal courts dealing with federal agencies, part of the Chevron doctrine63 calls for lexical ordering with a logical hinge that turns on clarity: Federal judges are not supposed to defer to agency interpretations when the meaning of a federal statute is adequately clear on the issue at hand.64 Almost surely the doctrine means to stop the agency’s position from influencing the judiciary’s evaluation of statutory clarity.65 Otherwise deference would creep into the decision of whether to defer. There are other parts to the doctrine—only some agency interpretations are eligible for judicial deference (Step Zero),66 and courts are apparently supposed to apply some kind of reasonableness test to agency positions when the statute is unclear about the issue at hand (Step Two).67 Nonetheless, the Chevron doctrine’s statutory clarity hinge in Step One is analytically comparable to the treatment of legislative history discussed above.

58 See, e.g., Alex Sheshunoff Mgmt. Servs., L.P. v. Johnson, 209 S.W.3d 644, 652 n.4 (Tex. 2006) (“If the text is unambiguous, we must take the Legislature at its word and not rummage around in legislative minutiae.”). Here “at its word” must refer to the particular statute at issue, not the Legislature’s code construction act.

59 See BankDirect Capital Finance, LLC v. Plasma Fab, LLC, 519 S.W.3d 76, 85 (Tex. 2017) (“[W]e have resolutely refused the Act’s entreaties to disregard plain language . . . . 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.”).

60 State v. Gaines, 206 P.3d 1042, 1051 (Ore. 2009) (reiterating that “what [legislative history] is worth is for the court to decide”).


64 See, e.g., Kingdomware Technologies, Inc. v. United States, 136 S.Ct. 1969, 1979 (2016); Chevron, 467 U.S. at 842 (“If the intent of Congress is clear, that is the end of the matter . . . .”).

65 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 Strauss, supra note 22, at 1164–65.


67 See Chevron, 467 U.S. at 843–45, 851, 865–66 (referring to a “permissible” or “reasonable” decision); Matthew C. Stephenson & Adrian Vermeule, Chevron Has Only One Step, 95 Va. L. Rev. 597, 599–603 (2009)
Chevron was handed down in 1984—during an unsettled transition period between American Trucking’s inclusiveness and the rise of textualism and ultimately lexical ordering at the Supreme Court in the 1990s. Chevron can be portrayed as displacing a multifactor test that was supposed to calibrate the degree of judicial deference to the persuasiveness of the agency position, and then aggregate whatever the agency’s position was worth with other interpretive sources. Fittingly, American Trucking itself cited not only legislative history but also an administrative agency’s position as a factor in its decision. Chevron was unanimous among the seven Justices participating, and we might wonder whether part of the doctrine’s attractiveness was the merger of plain meaning with a secondary consideration. In any event, Chevron was an intellectual forerunner to a broader set of lexical ordering moves at the Court.

State-level embrace of Chevron-style lexical ordering is somewhat less widespread but still common. Extreme outliers do exist. In the doctrine of approximately five western state courts, no deference or weight is afforded a home state agency’s interpretation of a home state statute. Additionally, Iowa courts demand that a state statute clearly delegate discretion to the state agency on the disputed issue before entertaining deference. 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.

See Manning, supra note 25, at 162–63; Thomas W. Merrill, Judicial Deference to Executive Precedent, 101 Yale L.J. 969, 972–73 (1992) (stating that, pre-Chevron, the Court deployed a multifactor approach and comparing canons of construction); Adrian Vermeule, Mead in the Trenches, 71 Geo. Wash. L. Rev. 347, 348 (2003) (“The Chevron opinion itself is best read as an attempt to simplify and clarify . . . .”).


At least three of those state courts have equivocated somewhat recently. See Molnar v. Fox, 301 P.3d 824, 831, 370 Mont. 238, 246–47 (Mont. 2013) (indicating that longstanding agency interpretations are influential); Bostwick Properties, Inc. v. Dep’t of Nat’l Res. & Conserv., 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. Liquor Control Comm’n, 810 N.W.2d 149, 162–63 (Neb. 2012) (acknowledging occasional court statements that department interpretation receive weight, especially after the state legislature fails to change the interpretation); Public Serv. Comm’n of Wyo. v. Qwest Corp., 299 P.3d 1176, 1182 (Wyo. 2013) (articulating a Chevron-like formulation).

See Myria Holdings Inc. v. Dep’t of Rev., 892 N.W.2d 343, 347 (Iowa 2017) (“[W]e only grant deference to an agency’s interpretations of law if the particular matter is clearly vested by statute in the agency’s discretion.”). Cf. IOWA CODE § 17A.19(11)(b)–(c) (enacted in 1998) (addressing deference but without the term “clearly”).

See, e.g., N.M. STAT. § 12-2A-20.84 (listing administrative construction in a top-tier of factors); Marbob Energy Corp. v. New Mexico Oil Conserv. Comm’n, 206 P.3d 135, 139, 146 N.M. 24, 28 (N.M. 2009) (stating that the court is merely “less likely to defer” when the statute is clear); cf. Revised Model State Administrative
Some states have deference doctrines without a step that is easily translated into an independent judicial assessment of statutory clarity.\textsuperscript{73}

But most state-level doctrine is in accord with \textit{Chevron} on the priority of clear statutory meaning. In recent years, approximately forty state courts have indicated that they support a lexically superior step in which judges are supposed to independently determine whether a statute is sufficiently clear or unambiguous to foreclose giving deference or weight to an agency interpretation.\textsuperscript{74} In Massachusetts, for example, courts tell us that they review agency interpretations de novo, but they also depict an analytical structure that largely tracks \textit{Chevron}, asking first “whether the Legislature has spoken with certainty on the topic in question.”\textsuperscript{75}

* * *

Courts have not all settled on precisely which sources belong in which tiers, and struggles over the tiers are to be expected where lexical ordering is endorsed. For federal judges, the tier above legislative history no doubt includes ordinary word meanings, words surrounding a term in dispute, and inferences from text in the rest of the same enactment.\textsuperscript{76} But there is room for debate over how inclusive the top tier ought to be.\textsuperscript{77} And the hard questions of how to appropriately weight and interact various sources within the same tier will not go away. The challenges are likewise when an agency interpretation is available. There has been disagreement

---

\textit{Procedure Act}, 15 U.L.A. 1, 96–97, § 508 cmt. (2010) (“This section is not intended to preclude courts from according deference to agency interpretations of law, where such deference is appropriate.”).

\textsuperscript{73} See, \textit{e.g.}, Murray v. Utah Labor Comm’n, 308 P.3d 461, 469, 473, 2013 UT 38, ¶ 40 (2013) (distinguishing no deference for pure questions of law from the possibility of some deference for mixed questions of law and fact).

\textsuperscript{74} About thirty state supreme courts have, in my view, plainly taken this pro-lexical ordering approach to agency deference in recent years. See, \textit{e.g.}, Nielsen Co. (US), LLC v. County Bd. of Arlington Cnty., 767 S.E.2d 1, 4–5, 289 Va. 79, 88 (2015) (“[A]bsent ambiguity, the plain language controls and the agency’s interpretation is afforded no weight beyond that of a typical litigant.”). In about ten other states, the supreme court’s current position is not quite as easy to code but notable indications of such lexical-ordering preferences exist. See, \textit{e.g.}, Motor Vehicle Admin. v. Krafli, 158 A.3d 539, 547, 452 Md. 589, 603 (2017) (stating that courts “may accord some weight” to an agency's statutory interpretation without advertting to statutory clarity); Md. Ins. Comm’r v. Cent. Acceptance Corp., 33 A.3d 949, 958; 424 Md. 1, 16 (2011) (stating years earlier that “when the language of the statute is clear and unambiguous, no deference is due the administrative interpretation”). A spreadsheet of authorities is available from the author. See supra note 41; see also Bernard W. Bell, \textit{The Model APA and the Scope of Judicial Review: Importing Chevron into State Administrative Law}, 20 WIDENER L.J. 801, 818–19 (2011) (collecting sources, and citing Delaware and Michigan as states that have rejected \textit{Chevron}).

\textsuperscript{75} Biogen IDEC MA, Inc. v. Treasurer & Receiver Gen., 908 N.E.2d 740, 750 (Mass. 2009) (citing \textit{Chevron}) (citation and internal quotation marks omitted); see \textit{id}. (stating that, at the first step, the court uses “conventional tools of statutory interpretation”); see also Peterborough Oil Co. v. Dep’t of Envt’l Prot., 50 N.E.3d 827, 832 (Mass. 2016) (similar).

\textsuperscript{76} See King v. Burwell, 135 S.Ct. 2480, 2489–90 (2015) (relying on these three resources to conclude that the Affordable Care Act was unclear on the question whether subsidies were available for people purchasing health insurance on federally operated exchanges); see also Robinson v. Shell Oil Co., 519 U.S. 337, 341 (1997). Also, absurdity remains in the top tier. See, \textit{e.g.}, Arlington Cent. Sch. Dist. Bd. of Ed. v. Murphy, 548 U.S. 291, 296 (2006) (“When the statutory ‘language is plain, the sole function of the courts—at least where the disposition required by the text is not absurd—is to enforce it according to its terms.’”) (citation omitted).

\textsuperscript{77} Compare Yates United States, 135 S. Ct. 1074, 1081–83 (2015) (plurality) (relying in part on a statutory title to support a narrow understanding of a criminal statute, although citing precedent for the use of titles to clear up otherwise unclear statutes), with \textit{id}. at 1090 (Alito, J., concurring) (stating that statutory titles can be useful in resolving doubt about statutory meaning), and \textit{id}. at 1094 (Kagan, J., dissenting) (asserting that statutory titles cannot limit plain meaning in operative provisions).
over whether legislative history should be considered at *Chevron* Step One, or whether the demotion of legislative history to a secondary tier in other cases means that legislative history should drop out of Step One.\(^{78}\) Moreover, lawyers and judges may and to do argue heavily from judicial precedent in statutory cases. As a matter of practice, perhaps precedent rests in the very top tier of sources.\(^{79}\)

Furthermore, statewide or nationwide lexical ordering efforts might have different influences on different judges. Not every judge will follow the decision methods endorsed by their colleagues,\(^{80}\) and, regardless, the “clarity” tests built into today’s lexically ordered statutory interpretation might well leave significant room for each judge to maneuver.\(^{81}\) Plus, majority opinions may backslide on what had appeared to be the orthodox approach. The opinion for the Court in *King v. Burwell* quickly—perhaps bashfully—referenced congressional committee work to develop a background understanding of what the Affordable Care Act was trying to achieve and how.\(^{82}\) Prescribed interpretive methodologies are not fully stable across time anyway. All that said, working majorities of judges have emerged across the country to declare that they have pushed down legislative history and administrative interpretations into a lower tier of sources. This development requires our attention.

### II. Trade-Offs in Theory

In most judiciaries today, lexical ordering for statutory cases is orthodoxy. But not every judge can be comfortable with this logical structure; many judges prefer methods that flatly exclude certain lower-tier sources or that instead include those sources in a single mix of considerations. Regardless, decision practices in every jurisdiction are open to change over time, by some combination of people committed to reform. State and federal judges seem likely to take the lead in formulating approved interpretive methods for statutory cases, as they have in the recent past.

The deep and open questions are whether (or when) lexically ordered statutory interpretation is justified given its peculiar trade-offs—and whether (or when) judges can implement these instructions successfully when deciding cases, not just when writing opinions. This Part isolates downsides and upsides of lexically ordered interpretation, as a matter of theory. The discussion begins by assuming away decision costs and the risks of indecision, and then

---

\(^{78}\) See, e.g., New Mexico v. Dep’t of Interior, 854 F.3d 1207, 1227 (10th Cir. 2017) (assuming without deciding that legislative history may be considered at *Chevron* Step One when statutory text and structure seem clear); 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 not be considered at *Chevron* Step One, but rather at Step Two).

\(^{79}\) See *Kimble v. Marvel Entertainment*, LLC, 135 S. Ct. 2401, 2409 (2015) (reiterating that stare decisis is supposed to have “enhanced force” for cases interpreting statutes).

\(^{80}\) Federal judges seem skittish about methodological stare decisis. See Evan J. Criddle & Glen Staszewski, *Against Methodological Stare Decisis*, 102 GEO. L.J. 1573, 1576–77 (2014) (asserting that “federal courts generally do not give stare decisis effect to their methodological decisions in statutory interpretation cases”) (footnote and emphasis omitted); cf. Sydney Foster, *Should Courts Give Stare Decisis Effect to Statutory Interpretation Methodology?*, 96 GEO. L.J. 1863, 1872–77 (2008) (concluding that the Supreme Court has not expressly rejected methodological stare decisis in statutory cases, but that its opinions do not appear to follow such doctrine).

\(^{81}\) See infra Part III.B.

\(^{82}\) See *King*, 135 S.Ct. at 2486 (citing a Senate committee hearing).
introduces these phenomena to give lexical ordering a better defense. The next Part makes the analysis more realistic by exploring implementation problems.\textsuperscript{83}

\textit{A. Ideal Decision Situations}

Suppose first and unrealistically that interpretation is costless and that interpreters face no problem reaching decision. Then lexical ordering becomes not only odd but troubling and puzzling. Apart from alphabetizing items for easy reference, people usually have far better ways of gathering knowledge, reaching conclusions, and pursuing valued goals than the lexical ordering of considerations. In ideal decision situations, it usually makes no sense to lexically order information that is relevant to the decision at hand, and it makes no sense to consider irrelevant information at all.

The core risk is valuable information loss. As long as the lower-tier information is relevant to the decision, lexical ordering means that decision makers lose whatever value it has for the fraction of decisions in which the top-tier information alone is deemed adequate. Thus if decision makers conclude that top-tier information is sufficiently clear about 50% of the time but find lower-tier information useful about 50% of the time, then about 25% of all decisions will suffer the exclusion of useful information.\textsuperscript{84} In this quarter of the docket, lower-tier information has no power to make the call closer than it first seemed, or alter the formulation of governing doctrine, or even flip the final judgment in another direction.

True, lower-tier information presumably is less valuable than top-tier information. Perhaps decision makers can carefully choose (categories of) lower-tier information such that the probability of it being useful is far lower than 50%. As well, lowering the probability of stopping with the top-tier sources below 50%, such as by narrowing what counts as “clear” meaning, will boost the chances of reaching lower-tier sources and hence decrease the risks of ignoring useful information. Decision makers also might be sure to load up the top tier with as many probably useful (categories of) sources as possible, to ensure the consideration of this information. But making any of the foregoing moves will merely lower the probability that lower-tier information could affect decisions or be ignored if it would. Why tolerate \textit{any} such risks? When the lower-tier information turns out to make no difference, no harm is done by considering it—not on our initial assumptions. At this stage, the losses from lexical ordering are deadweight.

All of these simple thoughts are consistent with more sophisticated normative decision theory. One straightforward position on information in ideal decision situations is that “more is better.”\textsuperscript{85} Assuming the relevance of the information, increasing the amount of information considered should increase the chance of an accurate, reliable, and otherwise high-quality decisions. More formally, the principle of total evidence for rational actors recommends the use

\textsuperscript{83} The goal here is to advertise difficult, crucial, and underappreciated issues for lexical ordering in statutory cases, rather than a definitive conclusion for every statutory case. Having done little work on the deep questions surrounding lexical ordering, we have to take initial steps that leave some questions open.

\textsuperscript{84} 0.5 * 0.5 = 0.25. This assumes that the probability that the lower-tier information is useful is independent of the top-tier information seeming adequate. These numbers are loosely based on results from the survey described below in Part III.B.

of all available evidence when estimating the likelihood of various outcomes. In 1967, the British mathematician Jack Good offered a formal proof that, for those trying to maximize expected utility, “it pays to take into account further evidence”—if one ignores the cost of collecting and using the information. One can make the same claim about any information relevant to understanding and applying statutes.

Inclusive theoretical approaches to information, which leave no apparent room for lexical ordering, do leave room for different weights or persuasive values for different considerations. Rational choice and expected-utility maximization theorists have developed weighted-additive models for integrating a large number of decision factors, giving them weights and assigning them probabilities so that all relevant considerations may be accounted for. Weights may follow the estimated reliability of particular interpretive sources or broader source categories. Similar in this regard is Bayesian updating. A decision maker’s beliefs about the likelihood of a proposition can be adjusted based on new information, integrating her priors with the implications of incoming information. Any number of sources relevant to statutory cases can be assigned various weights: ordinary meaning of statutory text can be assigned very heavy influence, for instance, while statutory titles receive very little but not zero weight.

At the same time, thoughtful decision makers might flatly exclude troublesome considerations. In life and law, people try to avoid relying at all on a set of disfavored considerations for particular decisions, even when the information is cheap to process. For instance, however people might choose their friends, most people want to ensure that in most circumstances job applicants are not hired solely because of their religion or race or sex. A combination of moral, ethical, and legal commitments steer people away from reliance on particular reasons for making particular decisions. Lexical ordering has nothing really to offer such principled commitments—including the opposition to legislative history as a bad substitute for statutory text, or to shifting interpretive power from judges to agency officials.

Nor is lexical ordering a conventional evidentiary form of “conditional relevancy.” Information sometimes is deemed irrelevant in the sense of having no practical value unless some proposition is at least possibly true or false. Conditional relevancy is part of the law of evidence, and it has companions in ordinary logic. A rational employer may conclude that job

87 Id.
88 See Samaha, supra note 17, at 457.
89 See id. at 456–57.
90 On the usefulness of ordinary word meaning as a coordination device, without necessarily defending lexical ordering, see David A. Strauss, Why Plain Meaning?, 72 NOTRE DAME L. REV. 1565, 1565–66 (1997), and Frederick Schauer, Statutory Construction and the Coordinating Function of Plain Meaning, 1990 SUP. CT. REV. 231. Compare also the conventional rankings of influence for various subcategories of legislative history. See George A. Costello, Average Voting Members and Other “Benign Fictions”: The Relative Reliability of Committee Reports, Floor Debates, and Other Sources of Legislative History, 1990 DUKE L.J. 39, 41–42. For similar reasons, lexical ordering is not about setting presumptions, however strong. Presumptions remain within the overall mix of considerations and might be overcome.
91 See Fed R. EVID. 104(b) (2011) (“When the relevance of evidence depends on whether a fact exists, proof must be introduced sufficient to support a finding that the fact does exist.”); see also id. Adv. Comm. Note 1975 (“[W]hen a spoken statement is relied upon to prove notice to X, it is without probative value unless X heard it.”). The Rule does not spell out the dimensions of the idea, however, which can be difficult to separate from the Rules’ broad notion of “relevance.” See, e.g., Dale A. Nance, Conditional Relevance Reinterpreted, 70
applications should not be evaluated unless the employer has adequate revenue to hire; a rational judge may conclude that no work on penalties should be done unless the defendant’s guilt is established; a rational empiricist may conclude that there is no use studying the effects of lexical ordering in law unless law actually calls for lexical ordering. The idea is that certain information is just not helpful unless a specified condition is met or assumed.

This idea is, however, obviously different from claiming that certain information is not important enough to consider unless other resources wash out. Nobody is arguing that legislative history or agency interpretations are not “relevant” in an evidentiary sense to the questions of how to understand and apply the statutes at issue in court cases unless statutory text is otherwise unclear. Either way, legislative history and agency interpretations are assertedly “relevant” to these litigated questions. It’s just that many judges take the position that otherwise clear statutory text should stop the analysis even if lower tier sources indicate a different statutory meaning and a different judgment.

Thus lexical ordering is not about the familiar rational decision tasks of excluding, weighting, or adding up inferences from information, nor is it a conventional form of conditional relevance in evidence. Mathematically sophisticated decision-making models are not used much by ordinary people making ordinary choices, of course, but some of the foregoing ideas are common and intuitive. The upshot is similar regardless: Decision makers who are liberated from decision costs normally should include or exclude, not lexically order.

It still might be possible to justify lexical ordering in a principled fashion without resort to decision costs and indecision risks, but an adequate theoretical justification is challenging. A good justification for lexical ordering cannot rest with the conclusion that certain sources are “the best” evidence of legislative intent or statutory meaning. However true, this conclusion is no defense for the multi-tiered structure at issue. Identifying the best interpretive source is no more supportive of lexical ordering than of assigning heavy weight to that source within an aggregated mix of considerations.


92 I write “may conclude” because, among other considerations, developing information on the appropriate and likely penalty might facilitate plea-bargaining or settlement negotiations in the civil context. A rational judge might believe that such non-trial dispute resolution is valuable.

93 See Baude & Doerfler, supra note 9, at 2 (asserting that the relevance of information to interpretation normally should not vary depending on whether a legal text is clear).

94 John Rawls famously laid out an intricate template for testing the justice of political systems that included lexical superiority for some values over others. See RAWLS, supra note 19, at 42–43. Under his template, 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 based 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 is 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. In this Article, we confront lexical ordering as a type of compromise for judges deciding statutory cases.

95 E.g., Scher v. Burke, 395 P.3d 680, 688 (Cal. 2017) (observing that “[t]he Legislature’s language is the best indicator of its intent” and indicating that legislative history rests in a lexically inferior tier); People v. Bradford, 50 N.E.3d 1112, 1115 (Ill. 2016) (similar); Adams v. State, 960 N.E.2d 793, 798 (Ind. 2012) (similar).

96 See Kimmel v. State, 29 N.Y.3d 386, 397 (2017) (observing that “the plain language of the statute provides the best evidence of legislative intent” without making legislative history lexically inferior).
The proponent’s task is to show that certain top-tier considerations are so comparatively important that lower-tier considerations should have zero effect on results no matter how persuasive the latter may be alone, unless the former are basically in equipoise—and yet the lower-tier considerations must not be so insignificant that they ought to be excluded entirely from the analysis. This analytical needle might be threaded with informational sources that are morally fraught to the point of being nearly illegitimate considerations, shunted aside until other resources are truly exhausted. 97 Without extra normative motivation such as the need to economize on decision costs or to reach decisive judgments, this kind of principled justification for lexical ordering must be rare. But it is not inconceivable that an argument could be flushed out for valuing, say, the source category of legislative history in this way.

For statutory interpretation, John Manning began the task of justifying the demotion to lexically inferior status of purpose informed by legislative history. He did not argue judicial decision costs.98 Manning contended that one plausible version of purposivism would rule out the old notion of “the spirit of the law” trumping its text and instead follow the notion that clear statutory text should trump evidence of background purposes collected from legislative history.99 This theory seeks to respect congressional choices between specific rules that rather tightly constrain the discretion of statutory interpreters and vague standards that effectively delegate some implementation authority, while “enabling” Congress to make such choices via statutory text.100 Manning rightly emphasized the adoption of clear statutory text’s primacy by judges who do not adopt hardline textualism, and he was right to explore such lexical ordering as a matter of theory beyond pragmatic compromise alone.101

Manning could be right and his work is thoughtful, but the case for lexically ordered interpretation remains weak. Much of Manning’s analysis is effective against purpose trumping statutory text understood with a limited set of sources, which might disable Congress from using statutory text to make implemental choices. 102 But the Holy Trinity-style position under attack is itself a version of lexical ordering. Knocking out the lexical superiority of purpose or spirit does not support the lexical superiority of any other source. After all, the options for judges are not only about which sources to lexically order, but whether to lexically order at all. Worth heavy emphasis here is that sources thought to be highly valuable (such as ordinary meaning of statutory text or 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

---

98 See Manning, *supra* note 25, at 147–65 (stressing institutional settlement and constitutional law).
99 See id. at 115–16, 129–32.
100 See id. at 147–48, 163–65.
101 Recall Justice Kagan’s declared position that legislative history should not be used to create ambiguity “for those who take [legislative history] into account.” Milner v. Dep’t of Navy, 562 U.S. 562, 574 (2011); see also Honeycutt v. United States, 137 S. Ct. 1626, 1635 (2017) (“For those who find it relevant . . . .”). This appears to be a recognition that judges are exercising individualized choices over whether to consider legislative history at all, indicating that lexical ordering is not a compromise meant for every judge and perhaps not viewed as a compromise by the judges who endorse it.
102 Judges also must face the problem of determining 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 divide its labor). Manning’s analysis emphasizes constitutional constraints.
assigned extremely low weights—or entirely excluded from consideration.\footnote{At the end of the article, Manning refers to purpose “count[ing] for little” among nontextualist Justices. Manning, supra note 25, at 181. A very low weight for a source category will make that category relatively uninfluential to the decision and thus possibly not worth the effort. But this idea relies on decision costs.}

Lexically ordered interpretation should show better than these alternatives.

B. Non-Ideal Constraints

More hope for lexical ordering arrives when we add decision costs and indecision risks. The former is a standard consideration in information economics and institutional design theory, while the latter problem is less widely flagged. But both decision costs and the problem of stalemate provide rational reasons for decision makers to ignore some relevant information sometimes, which is a price that lexical ordering exacts. The challenge is identifying empirically convincing domains in which lexical ordering makes sense.\footnote{Lexically ordered interpretation sometimes has more than two tiers. In this discussion, I take up the trade-offs associated with a two-tier structure, which is the least complicated version of lexical ordering and therefore a sensible place to begin the analysis.}

1. Decision costs

Economizing on decision costs can be a reason to lexically order information. Rational decision making and sensible institutional design accounts for not only the quality of decisions and their consequences, but also the costs of reaching decision at various levels of quality. Often enough, we are confronted with unwelcome trade-offs in which decision costs and decision quality rise and fall together. Although saying so might be uncomfortable in the context of judicial decision making, where public officials offer final judgment in cases that affect the well-being of countless others, at some point increases in decision costs will not be worth the predicted gains in decision quality.\footnote{See, e.g., 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]; see also Strauss, supra note 90, at 1565–66 (discussing plain meaning rules and observing that “[s]ometimes it is more important that things be settled than that they be settled right.”); cf. Avery Wiener Katz, The Economics of Form and Substance in Contract Interpretation, 104 COLUM. L. REV. 496, 524 (2004) (discussing contract interpretation and observing that “a fuller or broader context can be purchased—but only at a cost of time and trouble, and of exacerbating certain incentive problems--so it pays to stop at some optimal point”).

Lexical ordering might resemble a “take the best, ignore the rest” heuristic, Gerd Gigerenzer & Daniel G. Goldstein, Reasoning the Fast and Frugal Way: Models of Bounded Rationality, 103 PSYCH. REV. 650, 653 (1996), in which people facing heavy information loads screen out options using one or two important considerations and leave the remaining options for more thorough evaluation, see 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 large and complex option sets). Initial screening may rule out the option that would have been found best, if only the decision maker had thoroughly evaluated it using every relevant metric. The heuristic might or might not be rational depending on decision costs, stakes, and other factors. Cf. Baude & Doerfler, supra note 9, at 551 (considering briefly a cost-efficiency justification, and distinguishing plain meaning rules from a heuristic in which some options are ruled out initially as clearly inadequate but without making a final decision on those grounds alone).}

Collecting and considering information always costs somebody something, but only sometimes yields useful insight for a decision.\footnote{Even a source category that judges always deem useful might be flatly excluded from consideration if the decision costs surrounding it are high enough. Interpretive sources for statutory cases are usually not so consistently valuable and extremely costly to process, though, so we may focus on more realistic trade-offs.} Which gives us a simple reason to want judges
to run some risks by sometimes ignoring a well-chosen set of lower-tier sources. We also have
the beginnings of a principle for placing sources in different tiers: To minimize the
informational value lost in the pursuit of cheaper decision making, the most valuable sources
should be placed in the top tier and the least valuable in the lowest tier, all else equal.\footnote{See Samaha, Tiebreakers, supra note 9, at 1699–1700; cf. Louis Kaplow, Optimal Multistage Adjudication, _ J.L. & ECON. ORG._ (forthcoming 2017) (“[F]or some legal systems, the structuring of the stages [of adjudication] is itself a decision variable, so one could analyze, for example, whether consecutive stages should be combined (balancing cost reduction due to economies of scope in information collection against forgone option value) and how stages should optimally be ordered (information with a high ratio of diagnosticity to cost would optimally be gathered first).”). One simplifying assumption here is that all sources are equally costly to collect and consider, and those costs are independent rather than affected by the other sources in the same tier.} Eventually we will have to compare alternatives such as simply excluding low-value information or including all relevant information in a single mix. But this is a start.

Extending our numerical illustration,\footnote{See supra note 84 and accompanying text.} suppose that legislative history has a 50% chance of being useful to the decision,\footnote{The broader the definition of “useful,” the less able judges will be in reducing this estimated probability. If “useful” merely means entering judgment for the right party, and not the correct articulation of governing legal doctrine or the correct sense of how close the case is, judges will have an easier time the screening out non-useful sources.} but considering it will always cost something. Hopefully the expected decision benefit will meet or exceed the expected decision cost to justify the resort to legislative history.\footnote{In (off-putting) math terms, $p^b \cdot b \geq p^c \cdot c$, where $p^b$ is the probability of a benefit, $b$ is the benefit, $p^c$ is the probability of a cost, and $c$ is the cost. The probability of cost is 1.0 in this hypothetical so the term drops out.} Perhaps the chance of finding valuable information is too high for us to support judges overlooking legislative history in any fraction of cases—although everything now depends on how we value the benefits and costs, which we will revisit below. With low enough decision costs, though, judges should always consider legislative history even if it bears value only half the time. If we instead suppose that legislative history has only a 1% chance of being useful, we might support a wholesale exclusionary rule. Modest per case decision costs could swamp the low probability of finding useful information—although, again, everything depends on our valuation of benefits and costs in the context of judges deciding statutory cases.

Plausible space for lexical ordering might open up with variation across cases, joined with diagnostic potential for a lexical ordering rule. Suppose that legislative history is useful in 25% of all statutory cases, but that this probability varies partly based on the persuasiveness of higher-valued sources. In one type of case, inferences drawn from source categories that are expected to yield high informational value are so strong that even a crystal-clear inference from legislative history would be very unlikely to make a difference to the decision. In a second type of case, inferences drawn from higher-valued sources are so weak or contradictory that even a weak inference from legislative history becomes likely to make a difference. The first case type is like a 1% chance of value hypothetical, while the second case type is like our 50% chance of value hypothetical. Perhaps the lexical ordering rule for resorting to second-tier sources will separate the two case types.

It is not obvious how to formulate such a rule. But vaguely asking judges to determine whether a statute is otherwise “clear” might be operationalized as a test for low probability as opposed to high probability of usefulness in a second-tier source. When a statute is said to be “clear” as applied to the instant case, judges might mean that the top-tier sources are sufficiently
persuasive that there is little chance of a second-tier source mattering (our first case type). When the statute is said to be “unclear,” judges might mean that a second-tier source is likely to matter (our second case type). If so—and if judges make these determinations cheaply and reliably—then judges might well do better for us than wholesale rules of exclusion or inclusion. The direct decision costs of considering legislative history will be saved in cases where there is only a 1% chance of usefulness, but expended when the chance rises to 50%. Thus we can think about well-constructed lexical ordering as prioritizing information such that lower-tier sources are less often used despite their potential value, but used when most likely to be useful. It is a compromise that alternates between risky exclusion and costly inclusion, with some probability of either in each case.111

The trade-offs are becoming clearer but so are difficult valuation issues. Conclusions depend on variables about which we are basically speculating, such as the probability that legislative history will be “useful” across a range of cases, without examining it. Additional valuation questions are worth adding, too: How much is a judge’s time worth, exactly? How much is a fully resourced judicial decision worth, exactly? How much less valuable is a decision that relies on less than all relevant information?

We have yet to add the predicted upstream incentive effects on parties, lawyers, and legislatures from judges choosing one method over another. For instance, theoretically parties are more likely to pay lawyers to collect, argue, and even push legislators and staffers to generate favorable legislative history if judges always consider it, somewhat less likely to do so if it is lexically ordered into a lower tier, and less likely still if judges say they never consider it. A complete social accounting would include those factors as well.112 But what are the numbers? Is the middle ground of lexical ordering sufficiently distinctive to alter upstream drafting practices in legislatures?113 Nor have we considered how judges are likely to behave under each alternative official method, including how hard they will try to follow instructions. The behavioral effects might end up being minimal or nonexistent—and the incentive effects depend on judges holding firm over time to a given version of lexical ordering.114 Estimating, valuing, and aggregating all relevant trade-offs is a daunting task.115

111 Note that the analysis thus far omits the risk of judges erring in their understanding of legislative history.

112 See Kaplow, Optimal Multistage Adjudication, supra note 107, at ___ (concentrating on ex ante behavioral effects including information gathering, thereby adding complexity to standard decision analysis involving the costs and benefits of acquiring additional information before making the ultimate decision); Louis Kaplow, Multistage Adjudication, 126 HARV. L. REV. 1179, 1187 (2013) (similar). In multistage adjudication, the additional information is either not gathered or not necessarily presented to the decision maker in earlier stages. Statutory interpretation is normally not designed that way; usually lawyers submit briefs and make arguments on lower- and higher-tier sources at the same time. The decision architecture could be redesigned, at some cost.

113 As a logical matter, we should be extremely skeptical that this type of middle ground will have much different influence on drafting practices compared to judges telling us that they will consider the totality of valid and relevant interpretive sources for whatever they are worth. Furthermore, we should have little hope of detecting empirically a difference in the legislative behavior where judges adopt this interpretive approach as opposed to others. Because we lack excellent empirical evidence, readers are free to make their own evaluation here. But it is difficult to see how this kind of incentives-based defense can be persuasive for lexically ordered interpretation.

114 Judges probably have difficulty coordinating for sustained periods of time around an interpretive method that is at all controversial.

115 Accord Kaplow, Multistage Adjudication, supra note 107, at 1187 (“[O]ptimal system design and operation in
Impressions are what we have. And a problem for lexical ordering is that the avoided decision costs do not seem very high at this point. In terms of direct costs, usually we are talking about judges evaluating the legislative history that lawyers point at. Sometimes the available source material is quite limited.\textsuperscript{116} Yes, occasionally judges and their staff expend many hours gathering legislative history omitted by lawyers, understanding the debates and amendments, and contextualizing each bit of history into a meaningful whole.\textsuperscript{117} But each effort contributes to the resolution and prevention of subsequent disputes. The research does not have to be done from scratch every time a given statutory provision appears in litigation, and completed research detailed in a court opinion can help future cases settle. Moreover, today’s lexical ordering in statutory cases adds a “clarity” test to regulate the turn to lower-tier sources. The test is itself a decision-cost generating complication, which could be avoided by flatly including or excluding lower-tier sources.

The decision-cost justification for lexically ordering agency deference seems worse. The direct decision costs of judges turning to an agency position are bound to be even lower on average than collecting and considering legislative history. The agency interpretation stands there, formulated, waiting to be adopted.\textsuperscript{118} Still, the direct costs to judges of \textit{Chevron} style doctrine depend on how thorough and novel any Step Two reasonableness testing becomes. A quick check for a reasonable interpretive choice by the agency, in accord with conventional interpretive arguments, will cost judges very little. Judges will have already 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, and therefore presents nearly no decision costs for judges to avoid with lexical ordering.\textsuperscript{119}

2. \textbf{Decisiveness}

There is another reason to run the risks of lexical ordering, however. Lexical ordering is one response to the problem of intolerable ties.\textsuperscript{120} Intolerable ties occur when two or more options are equal on a given metric, or equal as far as we can tell with any confidence, and yet one option must be selected over others. Many legal institutions must deal with this dilemma. Judges and other officials are asked to reach decisive conclusions on even the most difficult issues without seizing up. Perhaps these institutions should be redesigned to allow for confessions of uncertainty as an acceptable form of final answer, but many of us are convinced that official and decisive resolution of some range of disputes is socially important.

\begin{itemize}
\item any context depend on many empirical matters that are heretofore unexplored and would be difficult to assess.\textsuperscript{9}).
\item \textit{See supra} Part I.B.1.
\item Think about a supreme court with rich decisional resources, delving into the major question whether young people are protected from age discrimination under federal law. \textit{See} Gen. Dynamics Land Sys., Inc. v. Cline, 540 U.S. 581, 586–92 (2004) (offering a wide-ranging discussion of legislative history and some social history).
\item Of course we should add any unaccounted social costs associated with agencies producing these interpretations in the hope of getting deference. But often the agency will produce a litigating position regardless.
\item A rule of lenity (or severity) costs judges basically nothing to use when deployed only as a tiebreaker. If lenity became one factor in a larger mix, however, it would need an assigned weight.
\item The analysis in the following paragraphs draws from Samaha, \textit{Tiebreakers, supra} note 9, at 1085–1700, and Samaha, \textit{Looking Over a Crowd, supra} note 105, at 567–89.
\end{itemize}
Reserving an interpretive source in a lexically inferior tiebreaking position tends to drive down the chance of all sources canceling out. This finding rests on several simplifying yet not altogether unrealistic assumptions. Suppose that any given source may strongly or weakly support either side’s position in a case, or be unclear as to the proper interpretation. Suppose further that each source is equally likely to have any one of these five possible implications, and that the judge will simply add up the implications of all sources considered. 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 overall uncertainty. The judicial dispute resolution system will have stalled.

Using permutation tables and simulations, it can be shown that the chance of such overall uncertainty decreases as the number of sources considered increases. The allied intuition is that tossing large numbers of sources together, like tossing lots of coins, is not likely to yield an exactly equal number and strength of implications pointing in two different directions. Hence one tactic for reducing intolerable ties is to increase the number of sources considered, which will increase decision costs but hopefully increase decision quality as well as decisiveness. On the foregoing assumptions, moving from one to five sources reduces the probability of all implications canceling out into overall uncertainty from 20% to about 12%.

Another option is to demote one source into a lexically inferior tiebreaking position, such that the demoted source is only considered when the other sources cancel each other out. As it happens and on the same assumptions, such lexical ordering tends to drive down the chance of overall uncertainty faster, relative to an aggregated mix of the same number of sources. Moving from one to five sources reduces the probability of overall uncertainty from 20% to under 3%, when one of those sources is held aside for tiebreaking. Holding one source aside does tend to increase the chance of a tie among the remaining top-tier sources, but this increase is more than offset by the decisive power of the tiebreaker, which has a four in five chance of offering a now-conclusive inference in one direction or the other.

---

121 We might hope and expect that the interpretive sources relevant to any given case will not tend toward equipoise. But we are modeling litigation, which often though not always involves represented parties contesting legally hard cases.

122 See Samaha, Looking Over a Crowd, supra note 105, at 572–75.

123 See id. at 575.

124 To be clear, decision makers cannot always add information, sensibly or at all. See Samaha, Tiebreakers, supra note 9, at 1689 (listing limits on the tactic of adding sources to reduce ties). First, decision costs tend to rise if additional information is collected and considered. Second, decision quality might suffer at some point. People sometimes make more mistakes when the decision protocol becomes excessively complicated or loaded with information. See Samaha, Looking Over a Crowd, supra note 105, at 590–93. Third, the number of sources relevant, not to mention legally valid, has an upper bound. Relevant information will run out. People certainly can develop new understandings of legislative history, for instance, but the primary sources presumably are capped by what happened in the past.

125 See Samaha, Looking Over a Crowd, supra note 105, at 572–73 & figs. 1–2.

126 See id. at 1691.

127 See id. at 1693 fig. 6 (comparing the declining probabilities for overall ties when there is a tiebreaker and no tiebreaker). There can be no such tiebreaking decision structure with only one source. With two interpretive sources and one of them reserved for a lexically inferior tiebreaker, the probability of a tie is 4%—compared to 20% if the implications of both sources are simply added up. See id.
Furthermore, if judges choose a type of tiebreaker that always points in a direction without ever being unclear, then the risk of a tie will fall to zero. Every case will close using only valid and relevant sources.\footnote{See Samaha, \textit{Tiebreakers}, supra note 9, at 1694.} Thus while the decision-cost justification for lexical ordering was probably stronger for legislative history than for agency interpretations, now the situation seems to be reversed. Agency interpretations tend to be simpler tiebreakers than legislative history. Agency interpretations presumably close every case for which they are offered, as long as judges do not complicate matters with serious reasonableness testing, while legislative history often requires significant judicial analysis without a guaranteed implication in one direction or another. The rule of lenity is an even cleaner tiebreaker for criminal cases, by the way, being outstandingly quick and decisive once brought into view.

Like the decision-cost justification, however, the decisiveness justification for lexical ordering depends on difficult valuation problems. Even if our simplistic model is accurate enough, in some fraction of cases judges would gain decisiveness (and avoid decision costs) by forfeiting a more complete picture of the universe of valid sources.\footnote{Assuming that the candidate tiebreakers are not “double counted.” See \textit{id.} at 1696–98.} In this case set, judges would overestimate or underestimate the clarity of all source implications taken together, overlook the true uncertainty of aggregated source implications, or reach an outright incorrect case result.\footnote{See Samaha, \textit{Tiebreakers}, supra note 9, at 1695 (illustrating losses numerically). Randomization to break ties does not require such information losses, see \textit{id.} at 1689–91 (identifying upsides and limits to random tiebreakers), but randomization often is not available in legal institutions for the overt resolution of disputes, see Adam M. Samaha, \textit{Randomization in Adjudication}, 51 WM. & MARY L. REV. 1, 29–53 (2009) (examining differing attitudes toward random case assignment and random merits judgments); see also Guido Calabresi & Philip Bobbitt, \textit{Tragic Choices} 41–44 (1978) (critiquing lotteries and comparing them to other allocation rules).} In our simplified illustration, the probability of an incorrect result is 8% with two sources that are equally likely to show any of the five implications listed above and with one of those sources reserved for tiebreaking. With five such sources and one reserved for tiebreaking, the probability of an erroneous result falls to about 5%—but is not eliminated.\footnote{See \textit{id.} at 1695–96 figs. 8 & 9 (showing percentages of overlooked true ties in the single digits or thereabout). This risk is somewhat higher with a binary or unidirectional source reserved for tiebreaking. See \textit{id}.}

Under what conditions, if any, is this risk acceptable in a judiciary? Judges can further reduce the risk of incorrect results by choosing less powerful sources for the tiebreaking tier. A source that cannot implicate any result strongly is less likely to flip the result suggested by the other more powerful sources. Still, reserving a relevant source for tiebreaking will occasionally overlook true ties, had all sources been mixed together.\footnote{See \textit{id.} at 1694 n. 86, 1696 fig. 8. These error calculations exclude cases in which the lexically superior sources add up to zero, because the lexically inferior source would have been considered regardless of lexical ordering.} On our assumptions, lexical ordering purchases extra decisiveness by obscuring the reality of truly hard cases. Lexical ordering will often obscure precisely how close cases are—perhaps in a majority of cases\footnote{See \textit{id}.}—if not influence judicial formulations of governing doctrine for future cases. Whenever lower-tier sources are ignored yet bear an implication for statutory meaning, the judge’s interpretive resources are skewed in some sense.\footnote{An alternative is to lexically order but “double count” by including a source in both the top tier and the} How to evaluate these consequences is itself in doubt.
Now we can consolidate lessons, in theory and in general: *Excluding sources* will tend to minimize decision costs, but the excluded sources must be carefully chosen because exclusion will tend to reduce decision quality and decisiveness. *Including sources* will tend to have the opposite effects, maximizing decision 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 logical hinge between tiers is designed. If the tiers are sensibly populated and if the hinge quickly and effectively separates cases in which lower-tier sources are either less or more likely to be useful to decision, then lexical ordering will tend toward a middling range of decision costs, a middling range of decision quality, and the highest rate of decisiveness.

All of the foregoing assumes, however, that judges understand and execute their lexical ordering instructions. To the extent that implementation is imperfect, the trade-offs will change.

### III. IMPLEMENTATION PROBLEMS

Machines can lexically order information without difficulty. In fact, there is a patch of computer science scholarship on machine learning that deploys lexical ordering to develop efficient search strategies. But of course people often process information differently from machines, even under substantively identical instructions. The specifics of how human decision makers process conditionally relevant information are still emerging in behavioral research. Judges might be in a special class with special abilities. Still, judges seem not to have made special efforts to design their decision environments to enhance the success of lexical ordering. Although the officially endorsed approach to statutory cases may require judges to temporarily ignore lower-tier sources, that information remains readily available to them. The rules of advocacy are loose enough to allow any lawyer to emphasize any interpretive source thought attractive to judges, regardless of the source’s formally assigned tier. Theoretically, judges could design their chambers, briefing, bench memoranda, hearings, and conferences to withhold lower-tier sources until each judge has reached a conclusion on statutory clarity. But we lack evidence that judges take any of these potentially costly precautions. When they do not, what happens to lexical ordering in practice?

#### A. Prior Research

Ordinary people are not always able to ignore information on purpose. First of all, people are not always sufficiently motivated to ignore information that they find attractive or

---

repulsive. A special version of the problem involves evidence-suppression instructions that backfire, making people more likely to rely on information that they have been told to forget. More generally, exposure to information can influence how people process other information. These phenomena pose implementation challenges for decision makers who have easy access to information that is supposed to be reserved for contingencies.

An acute problem area for lexical ordering in statutory interpretation is the “clarity” test for ignoring lower-tier sources. The test might be too vague to prevent judges from accessing lower-tier sources based on the attractiveness of those sources. This concerns gains some support from a study by Ward Farnsworth, Dustin Guzior, and Anup Malani, who asked first-year law students about the ambiguity of statutory text in a number of vignettes. Students were less likely to respond that they thought that criminal statutes were ambiguous when the students’ policy preferences were relatively strong. The students apparently were not asked for their final judgments as a matter of law, however. We do not know how often they thought the statutes clearly conflicted with their stated policy preferences, or whether asking for a final judgment would influence other responses. Furthermore, the correlation with strong policy preferences tended to go away when the question about ambiguity was framed as whether ordinary readers would likely disagree about the better meaning. Perhaps a carefully framed clarity test can reduce the influence of extralegal considerations.


See id. at 1262.

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 more generally, see Daniel M. Wegner, Ironic Processes of Mental Control, 101 PSYCH’L REV. 34, 35 (1994) (underscoring conditions of mental capacity constraint). Cf. Corbin A. Cunningham & Howard E. Egeth, Taming the White Bear: Initial Costs and Eventual Benefits of Distractor Inhibition, 27 PSYCH’L SCI. 476, 484 (2016) (finding that response times improved, with practice, when test subjects were given one color of letter to ignore in their searches).

See Wistrich et al., supra note 136, at 1264–70. A related example appears in first-impressions research where evaluations sometimes are more heavily influenced by information that arises early. See Kurt A. Carlson, Margaret G. Meloy, & J. Edward Russo, Leader-Driven Primacy: Using Attribute Order to Affect Consumer Choice, 32 J. CONSUM. RES. 513, 513–15 (2006); Samaha, Starting with the Text, supra note 17, at 462 (collecting studies).

See generally Wistrich et al., supra note 136, 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); id. at 1278 (flagging group deliberation as a potentially helpful tool for mentally suppressing inadmissible evidence). Alexander Platt has suggested that a “staging methodology” might reduce coherence-based reasoning in statutory interpretation. See Alexander I. Platt, Debiasing Statutory Interpretation, 39 OHIO N.U. L. REV. 275, 278 (2012). But any bias towards coherence cannot be cured in this way unless judges can and do mentally bracket the doctrinally separate stages of information processing.


See id. at 266–68.

See id. at 260–62.

See id. at 271–72, 292.
We should hesitate to draw inferences about judicial behavior from student responses.\textsuperscript{145} Yet few experiments involve real judges dealing with legal interpretation. In fact, we lack much research on the ability of experts of any kind to ignore information.\textsuperscript{146}

In a recent experiment, however, Dan Kahan and colleagues report that judges and lawyers responded differently from law students and the general public.\textsuperscript{147} The authors constructed two vignettes in which a criminal statute was (hopefully) ambiguous. The identity of the characters were then manipulated in legally irrelevant ways: a littering case involved water dispensers left at the U.S.-Mexico border, either by immigrant-aid workers or by border-fence construction workers; a disclosure case involved a police officer giving confidential investigatory information about a job applicant, either to a center offering birth control and abortion information or to a center that counsels on abortion alternatives.\textsuperscript{148} In the littering case, 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.\textsuperscript{149} Qualitatively similar results were obtained for the disclosure case. In contrast, the law student and general public respondents showed more sensitivity to the manipulation, consistent the expected influence of their cultural worldviews.\textsuperscript{150} The experiment studied legally irrelevant case facts rather than potentially admissible interpretive sources, but the case facts were ideologically charged and the results suggest that judges are a special type of decision maker.

Judges achieved more mixed results in experiments by Chris Guthrie, Jeff Rachlinski, and Andrew Wistrich involving inadmissible evidence.\textsuperscript{151} In one experiment, trial judges received a description of an armed robbery of $200 at a convenience store and were asked to adjudicate guilt in a bench trial. Some judges were also informed that the defendant confessed

\begin{footnotesize}
\item[145] See id. at 272–73 (recognizing potential differences with actual judges). Note, too, that judges may differ. Different judges serving on different courts might have importantly different traits and decision environments.

\item[146] On experts influenced by extraneous considerations, consider the small-$n$ but fascinating study of fingerprint analysis in Itel E. Dror, David Charlton & Ailsa E. Péron, Contextual Information Renders Experts Vulnerable to Erroneous Identifications, 156 FORENSIC SCI. J. 74, 76 (2006). The study took place in the experts’ workplace, with collaborators asking the experts to give an opinion on a high-profile 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 that there was no match. Most of the experts who were given the pro-match 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.

\item[147] See Dan M. Kahan, David Hoffman, Danieli Evans, Neal Devins, Eugene Lucci & Katherine Cheng, “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).

\item[148] See id. at 380–83.

\item[149] See id. at 394, 396 (figures 5 & 6) (showing point estimates); id. at 397–402 (reporting results of regressions and Bayesian analysis to estimate relative likelihoods of various hypotheses). The surveys did not ask about the respondents’ immigration or abortion policy preferences, but the respondents’ cultural outlooks were estimated in accord with Cultural Cognition Worldview Scales. See id. at 376.

\item[150] See id. at 394, 396–402.

\item[151] See Wistrich et al., supra note 136, at 1279–81 (describing the respondents as a collection of Arizona trial judges, federal magistrates, federal district judges, and trial court judges from an unidentified urban jurisdiction); id. at 1258–59, 1286–1312 (reporting results indicating inability to ignore attorney-client privileged information, excluded victim sexual history, and so on).
\end{footnotesize}
during a police interrogation that persisted after the defendant asked for a lawyer, and those judges were asked to rule on a suppression motion. The judges who were given no information about a confession indicated only slightly less willingness to convict than the judges who ruled that the confession was not admissible. The judges seemed to have ignored the excluded confession while ruling on guilt, in this simple robbery vignette anyway.

Guthrie, Rachlinski, and Wistrich later revisited the excluded confession challenge, adding variations in the severity of charges and the police misconduct. The pattern of influence was somewhat complex, with a variable that interacted the charges and the misconduct reaching statistical significance ($p = 0.04$). “More severe police misconduct reduced the judges’ willingness to convict” compared to mild misconduct, “but only for a less serious crime.” When instead the defendant was charged with robbery plus murder, “the judges who had heard confessions, however obtained, were consistently more willing to convict.” Differences across particular conditions were not always significant. Isolating the robbery-only conditions and the robbery-plus-murder conditions seemed to yield no statistically significant differences from adding a confession with either mild or severe police misconduct. Combining some of these conditions, however, did yield significant results. And the numbers are provocative regardless.

152 See id. at 1320–21, 1344–45; id. at 1321 n. 278 (noting that only one judge in the suppression condition would have admitted the evidence).

153 See id. at 1321 (reporting a difference of 3% and stating that there were 51 judges in the control condition and 53 judges in the suppression condition); see also id. at 1316, 1342–43 (reporting results of an excluded-evidence experiment involving an officer smelling drugs and wanting to search a car trunk, in which exposure to the fruits of the search did not seem to significantly effect the judges’ determinations of probable cause).

154 See Jeffrey J. Rachlinski, Andrew J. Wistrich & Chris Guthrie, Altering Attention in Adjudication, 60 UCLA L. Rev. 1586, 1611, 1613 (2013) (noting that the respondents were 81 federal district judges, 44 federal magistrate judges, 101 Florida state trial judges, and 88 California state trial judges, with approximately 50 respondents in each of six survey conditions).

155 See id. at 1611–12 (explaining that all judges were given a description of a bank robbery of $520, with some judges told about an ensuing murder of a young mother).

156 See id. at 1611–12 (explaining that judges were given descriptions either with nothing about a confession, with a detailed written confession including facts not disclosed by the police after the police twice refused the defendant’s request for a lawyer, or with the same confession after nine hours of police threats and other misconduct). The description of severe police misconduct included denial of access to a restroom and the defendant “soil[ing] his clothing.” Id. at 1612.

157 See id. at 1614 & n. 106.

158 Id. at 1614.

159 Id. at 1614–15.

160 See id. at 1613–15 & tbl. 2. Although not significant, in the robbery-only conditions, the conviction percentage went up with a confession and mild police misconduct (13%) but not with severe misconduct; in the robbery-plus-murder conditions, the conviction percentage was highest, oddly, with severe police misconduct. See id. at 1613 tbl.2.

161 See id. at 1614 n. 98 (explaining that, when the robbery-only and robbery-plus-murder conditions were combined, the increase in conviction rates between no-confession and mild-misconduct conditions reached marginal statistical significance ($p = 0.07$)); cf. Stephan Landsman & Richard F. Rakos, A Preliminary Inquiry into the Effects of Potentially Biasing Information on Judges and Jurors in Civil Litigation, 12 Behav. Sci. & L. 113, ___ (1994) (concluding that judge respondents, like other adults, were not able to ignore evidence of a tort defendant’s subsequent remedial measures despite being told that the previously presiding judge had held the evidence inadmissible).
At least under certain conditions, then, judges probably can ignore evidence that they have seen and then excluded. And although the decision situation is not exactly the same, judges might well be able ignore lower-tier interpretive sources while they concentrate on higher-tier sources to assess statutory clarity and perhaps end the analysis there. Indeed, we might think that the legally irrelevant evidence in the experiments above is more emotionally and ideologically charged than the ordinary run of interpretive sources in statutory cases, making the latter cases easier to resolve without improper considerations. But this is good territory for additional empirical research.

B. Clarity Survey

One implementation challenge for lexically ordered statutory interpretation is the “clarity” hinge between tiers of sources. Several commentators and judges have wondered what, exactly, the clarity test means across the wide range of cases and whether it is too vague to fence out improper considerations. One version of this concern is that judges will infect their clarity assessments with extralegal policy preferences, finding statutory text sufficiently clear or not depending on which option will promote the judge’s personally preferred case result. Testing this theory requires insight into judicial policy preferences, which can be opaque or weak in any given case. Studies like the Kahan group’s make progress on this front using judge respondents, but certainly we have more work to do.

Another version of the concern is that judges will be influenced by lower-tier sources in assessing statutory clarity, regardless of any preferences for a particular case result. Assuming that the judge took no special precaution against learning about lower-tier sources, those sources might influence how the judge evaluates statutory clarity using top-tier sources. The influence might be exerted in several ways and for several reasons. Apart from any case-specific influence from the content of particular lower-tier sources, for instance, a judge’s varying attraction to categories of interpretive sources might influence the likelihood of the judge finding clarity in the top-tier sources. The type of judge who tends to find statutory clarity using top-tier sources might be the type of judge who tends to find lower-tier sources useless—and vice versa.

Recall the ongoing debates over legislative history. Some judges and observers want legislative history excluded and others want it included for whatever it’s worth, but lexical ordering requires judges to hold aside lower-tier sources unless the statute is otherwise unclear. Judges with relatively extreme views might soften unwelcome implications of lexical ordering by conceptualizing and applying the clarity test conveniently—in two different directions. Perhaps judges who are averse to relying on legislative history also tend to see clarity in statutory text. On the flipside, perhaps judges who are enthusiastic about relying on legislative history also tend not to see clarity in statutory text. Both groups of judges could effectively execute

---

162 In the excluded-evidence experiments described above, judges who ruled that the evidence was admissible were excluded from the analysis. This choice focused the analysis on the influence of already excluded evidence on ultimate decision about guilt. A distinct question is whether exposure to challenged evidence influences the upstream decision to exclude that evidence, when the content of the evidence is not legally relevant.

163 See supra notes 12–13 (collecting sources).

164 See, e.g., Farnsworth et al., supra note 141, at 290 (suggesting that ambiguity determinations are easily influenced by strong policy preferences); Kavanaugh, supra note 12, at 2143 (similar).

165 See, e.g., Randolph, supra note 13, at 76.

166 See infra Part III.C (detailing a vignette experiment).
lexical ordering instructions—probably paying some mental tax—and yet reach results that are little or no different from the patterns that would emerge with a simpler decision procedure.\textsuperscript{167} Clarity might be in the eye of the beholder,\textsuperscript{168} and in patterned ways at that.

A new survey provides only weak support for the suggested pattern, however. Approximately one-hundred appellate judges from around the country were asked to generalize about their experiences with statutory text and legislative history.\textsuperscript{169} 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 (with 1 indicating never and 7 indicating 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 clarity in statutory text is associated with the judge’s attraction to legislative history, then we might see the numerical responses on these two questions running together and even clumping toward the high and low ends of both scales.

Figure 1 below displays a scatterplot of the responses.\textsuperscript{170} Indeed there are judges who indicated that statutory text is usually vague while legislative history is usually useful, as well as a larger group who indicated that statutory text is usually clear while legislative history is not usually useful. But these groups do not dominate the sample. The combined fraction of judges who responded with numbers below 4 for both questions or above 4 for both questions comprise less than one-third of the sample. Overall the responses are dispersed. While the responses on textual vagueness cluster toward 3, 4, and 5, the responses on legislative history’s usefulness are fairly spread up and down the plot. There does appear to be a positive correlation between the numerical responses on the two questions, but the correlation is weak (0.09) and not statistically significant at conventional levels ($p = 0.38$).

These measures of judge experience are imperfect, of course. Not every judge necessarily had in mind the same conception of “legislative history.”\textsuperscript{171} And a judge’s answer to a general question about legislative history’s usefulness could be influenced by recent exposure to a particular example within the larger category, such as the committee reports referenced in the first of two case vignettes described below.\textsuperscript{172} Reliance on committee reports is a relatively controversial practice, and the vignette was intended to place the committee reports in tension with some other interpretive sources. These factors could put downward pressure on the treatment group’s response to the usefulness of legislative history question, although the most likely effect on the correlation of interest here is not immediately apparent.\textsuperscript{173} The treatment group’s mean response to the usefulness question (3.71) is about one-half point lower than the control group’s mean (4.24). This difference is not statistically significant at conventional levels

\textsuperscript{167} Regardless, someone must draft opinions showing allegiance to formal law’s lexical structure.

\textsuperscript{168} See supra note 12 (noting such assertions).

\textsuperscript{169} For additional details on the survey, see Part III.C.1 below (noting that most of the judge respondents were state judges who had served for four or fewer years in their current positions).

\textsuperscript{170} One judge wrote “3-4” in response to both questions. That survey is excluded from this Figure.

\textsuperscript{171} See supra Part I.B.1 (observing that available state-level legislative history varies across the country).

\textsuperscript{172} See infra Part III.C.1.

\textsuperscript{173} The correlation looks weaker in the treatment group. In the control group, the correlation between the two responses was 0.19 ($p = 0.20$). In the treatment group, the correlation was -0.03 ($p = 0.83$).
using a two-sided $t$ test ($p = 0.14$), but comes closer to the 95% level using a one-sided $t$ test for whether the treatment group’s mean is lower ($p = 0.07$). One might resist the aggregation of these two groups’ responses, as in Figure 1.

**FIGURE 1: SCATTERPLOT OF APPELLATE JUDGE SURVEY RESPONSES, 2016–2017 ($n = 95$)**

In a more general spirit of caution, self-reported generalizations at one point in time are not always reliable measures of experience over time. In addition, judges with more experience might respond differently from the surveyed judges. Nonetheless, the results are worth contemplating. Surely some judges are susceptible to adjusting their assessments of statutory clarity in light of their attraction to legislative history in general, and a simple survey is one method of identifying them. The size of this group is open to question, however, and it might be small—which would leave the trade-offs identified in Part II largely intact.

174 However, an ordered probit analysis fails to confirm a treatment effect on the usefulness response at the 95% level ($p = 0.238$, two-sided).

A different complication for interpreting these results involves the rules of lexical ordering: Perhaps some of the 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 because they usually conclude that the statutory text is clear. I attempted to avoid this problem with the wording of the question, which asks about “cases that involve legislative history,” but I might not have been successful. In any event, asking about the usefulness of legislative history in cases “involving” it does not necessarily elicit judge opinions on usefulness if legislative history were examined in all statutory cases—let alone if all statutory provisions were examined or randomly sampled.
C. Vignette Experiment

To better understand the interplay between tiers of interpretive sources, we need more evidence. Until now, we have lacked a direct study of how judges execute lexical ordering instructions. Below are results from a survey 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 total participants, 86 were attendees at two seminars for new appellate judges from around the country, and another 16 were court of appeals judges serving in a Midwestern state. Most of the judges had fewer than four years of experience in their current judicial positions, which was a requirement for participating in the two seminars as a student judge. Among respondents for whom data is available, 78 were state judges, 8 were federal judges, and 10 were military judges (the other 6 judges did not respond to this question). To protect their anonymity and to increase response rates, the judges were not asked about their policy preferences or demographic traits.

Each judge was given two hypothetical cases to decide via paper-and-pencil survey. The vignettes are reproduced in the Appendix. In the Trade Name case, one company claims that another company is unlawfully using the same name for a different drug. The first company is not yet selling its drug on the market, however, and the disputed issue is whether pre-sale safety testing counts as being “used in commerce openly” under the quoted statute. In the Election Law case, employees who are paid an hourly wage claim that their employer unlawfully refused to pay them wages 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 the quoted statute.
Both vignettes were organized similarly, with a short description of the disputing parties, a quotation from statutory text, and party arguments back and forth on whether the given statute covered the defendant’s conduct. 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 issue area. Either way, to provide a serious and useful test of lexical ordering instructions, the competing interpretations of the statute should not be so evenly matched that very few judges view the statute as clear, yet they should be close enough for a given lower-tier source to be at least potentially influential if considered by the judges.

A randomly assigned treatment group received the descriptions as above, plus additional information and lexical ordering instructions. In Trade Name, the treatment group ($n = 51$) was given on-point legislative history, whereas the control group ($n = 50$) was not. The treatment group was told that uncontradicted committee reports indicate that the debated statutory language was intended to include safety testing. The message from this source thus clearly favored the plaintiff-company’s position, although judges were not instructed on how much weight to assign this source. In Election Law, the treatment group ($n = 50$) was given an administrative agency interpretation that clearly favored the plaintiff-employees’ position, whereas the control group ($n = 51$) was given no such information. 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.”

All judges were asked about statutory clarity at the end of each vignette. The control group was asked, in both cases, “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 numbers on the scale (1 = totally unclear, 2 = fairly unclear, 3 = somewhat clear, 4 = fairly clear, 5 = totally clear). These clarity questions are extraneous, as a matter of doctrine, in the absence of lexical ordering and lower-tier sources. But the questions were posed to the control group in order to make a direct comparison with the treatment group. The treatment group’s parallel questions reiterated the lexical ordering instruction 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?” The same descriptive notations were given. All judges also were asked for a final judgment in each case and for a level of confidence about each of those two judgments.

The survey 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 clarity response of 3, 4, or 5 meant that the lower-tier source in question was supposed to be ignored as a matter of law, whereas a clarity response of 1 or 2 in Trade Name meant that the committee reports should be considered, and a clarity response of 1 or 2 in Election Law meant

---

One possibility is that awareness of home-state law would tend to reduce the influence of sources presented in the vignettes because the answer became foreordained, thereby biasing the results downward, but other possibilities are worth considering.
that the agency position should be deferred to. As well, a judge in the treatment group who assigned a clarity level of 3, 4, or 5 was prohibited as a matter of law from altering the final judgment based on the lower-tier source in question. Omitting lower-tier sources from the control group’s vignettes will help us draw inferences about the effect of including such information on lexical ordering efforts. The control group was exposed to neither legislative history nor an agency interpretation in either vignette, and their set of responses can be used as a baseline for testing the influence of such lower-tier sources on judges. 

2. Hypotheses

• Clarity effects. Following the lexical ordering instructions in our experiment required the treatment group to respond to the statutory clarity questions while ignoring lower-tier sources. If the lexical ordering instructions were followed, then the clarity responses in the treatment group should not differ from those in the control group. Neither the means nor the distributions of responses on the clarity scale should differ significantly. For one or more reasons, however, some respondents might be unable or unwilling to mentally bracket these lower-tier sources while assessing clarity. Exactly how exposure to lower-tier sources might influence clarity responses is a more difficult question, though. And a significant difference at any point on our five-point clarity scale would suggest a problem for lexical ordering. The primary and simplest hypothesis for a lexical ordering failure is that (Hypothesis 1) clarity responses will differ between treatment and control groups.

• Clarity-plus-judgment effects. Final judgments are of interest, too. Following the lexical ordering instructions demanded that final judgments be stable—in part. Judges who assigned clarity levels of 1 or 2 were legally free (or required) to follow the lower-tier sources on final judgment, but not judges who assigned clarity levels of 3 or higher. We can look for evidence that exposure to a lower-tier source influenced certain combinations of judgments and clarity levels. Our ability to test for such effects will, however, depend partly 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 to the extent that judges responded that a statute was at least somewhat clear, regardless of exposure to lower-tier sources, but switched final judgments because of the exposure. Thus another and not mutually exclusive hypothesis is that (Hypothesis 2, judgment-flipping) among judges who assign clarity levels \( \geq 3 \), the fraction of judges entering pro-plaintiff final judgments will differ between treatment and control groups. Note, however, that this effect of exposure to lower-

---

178 This initial experiment varied the two vignettes along more than one dimension. Trade Name was supposed to be less ideologically charged and the treatment group was exposed to legislative history, while Election Law was supposed to be more ideologically charged and the treatment group was exposed to an agency interpretation. In addition, the agency’s position for the Election Law treatment group arrived with a brief rationale. Additional experiments will have to be run to explore the different effects, if any, of these dimensions of variation.

179 There are several possible tests for differences in clarity responses, besides a comparison of means. Testing for different means will not pick up, for example, parallel movements near the top and bottom of the scale that are offset by movements near the middle of the scale.

180 See supra text accompany notes 136–140 (collecting some explanations for failure to ignore information). This experiment will not distinguish potential psychological mechanisms.

181 Another hypothesis along these lines—not yet tested here—is that the fraction of all judges in the treatment group who enter pro-defendant (pro-plaintiff) judgments at clarity level \( \geq 3 \) will differ from the companion fraction in
tier sources would be especially strong. Moreover, we might have difficulty identifying this particular judgment-flipping effect, which must be distinguished from a more complex combination of effects.\footnote{Theoretically, exposure to lower-tier sources might have two simultaneous effects: (1) one set of judges lower their clarity level below 3 and switch judgments from defendant to plaintiff, while (2) a second set of judges raise their clarity to 3 or higher but stick with a pro-plaintiff judgment. This combination might be confused with judgment-flipping at clarity level 3 or higher, which is not mutually exclusive with the foregoing combination of effects. Any of these effects would indicate a problem for lexical ordering, however.
In addition, theoretically, cross-cutting judgment-flipping effects could cancel out if some judges are highly attracted to and other judges backlash away from lower-tier sources. The chance of these strong cross-cutting effects depends partly on the distribution of judgments in the control group.}

Finally, we can begin to examine narrower theories for the convenient use of clarity tests.\footnote{See supra Part III.B.} One theory suggests that judges who are attracted to lower-tier sources—whether as a category of information or based on case-specific content—will make those sources legally relevant by migrating to clarity level 1 or 2. A companion theory suggests that judges who are instead repelled by lower-tier sources will make those sources legally irrelevant by migrating to clarity level 3 or higher. In our vignettes, the contents of the lower-tier sources support the plaintiffs, while some fraction of the control group will conclude that the top-tier sources support the defendants. We can look for evidence in support of two additional hypotheses—(Hypothesis 3, attraction) the treatment group will move away from pro-defendant judgments at clarity level \( \geq 3 \) and toward pro-plaintiff judgments at clarity level \( \leq 2 \), and/or (Hypothesis 4, backlash) the treatment group will move away from pro-defendant judgments at clarity level \( \leq 2 \) and toward pro-defendant judgments at clarity level \( \geq 3 \).

Testing these convenient-use hypotheses might be difficult. Apart from our lack of evidence for how respondents feel about these lower-tier source categories,\footnote{We do have judge responses on the general usefulness of legislative history.} or their case-specific content, we might not be able to isolate confounding influences. The attraction and backlash theories do predict opposite effects on pro-defendant judgments at clarity level \( \geq 3 \) but the theories are not mutually exclusive, meaning that attraction and backlash could be happening simultaneously while their separate effects remain obscure in that combination of responses. Fortunately, the attraction theory predicts an increase in pro-plaintiff judgments at clarity level \( \leq 2 \) while the backlash theory does not. However, such an increase could come from a corresponding decrease in pro-defendant judgments at clarity level \( \leq 2 \), which would not indicate a lexical ordering failure. We should be willing to entertain plausible alternative explanations for any 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 making final judgments when they believe that the statute is otherwise adequately clear. True, specific support for the plaintiff’s position in uncontradicted legislative history is a tempting source for some judges, but the arguably low policy stakes might help legal professionals execute the instructions. The most plausible prediction for Election Law seems even 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 to rely on agency support, thereby combining legal resources with extralegal moves.

3. Results

The results are provocatively mixed. We see consistent warning signs that judges had trouble complying with the lexical ordering instructions in the Election Law case, but no such consistent signs in the Trade Name case.

- Clarity effects. As for Hypothesis 1, one option is to compare means. In Trade Name, the difference in the mean clarity responses between the control group (3.14) and the treatment group (3.35) is approximately 0.21 ($p = 0.27$, two-sided $t$-test). This result is at most mildly suggestive of a possible effect of exposure to the legislative history and does not reach conventional levels of statistical significance. In Election Law, the difference in means between control (2.94) and treatment (3.50) is approximately 0.56 and highly significant ($p = 0.006$, two-sided $t$-test). Exposure to the agency’s position seems to have influenced some judges’ clarity responses.

Another way of testing for a clarity effect is a Mann-Whitney rank sum test, which compares distributions without assuming normality and which is appropriate for an ordinal noncontinuous scale.185 Using this test, the headline is the same. In Trade Name, the difference in the distribution of responses on statutory clarity between treatment and control does not reach conventional levels of statistical significance ($p = 0.20$, two-sided Mann-Whitney test). In Election Law, however, the difference in distributions is highly significant ($p = 0.004$, two-sided Mann-Whitney test). Again the Election Law results are the more damaging for the prospects of successful lexical ordering.

Ordered probit models deliver the same headline, too. Ordered probit also is appropriate when the dependent variable is ordinal but not continuous,186 and these models allow us to control for a few additional independent variables for which we have data.187 Tables 1 and 2 report the results using six different specifications. The dependent variable is always the statutory clarity response on our five-point scale. Model 1 includes only a binary variable to indicate whether the judge was in the treatment group. Models 2 and 3 add variables representing the judges’ responses on the usefulness of legislative history in general and the vagueness of statutory text in general. Model 4 adds controls for whether the judge was part of the Midwestern state cohort, which comprised a more-experienced batch of judges from a single state, along with whether the judge served on a federal or military court. Models 5 and 6 introduce binary variables indicating whether the judge responded with a 6 or 7 to the question about the usefulness of legislative history (“LH useful high”) or a 1 or 2 (“LH useful low”).

In Trade Name, neither treatment nor any of the other independent variables are statistically significant at conventional levels in any of our models. In Election Law, the coefficient for treatment is positive and highly statistically significant ($p < 0.01$) across all models, while none of the other independent variables reach conventional levels of statistical

---

185 See SCOTT LONG, REGRESSION MODELS FOR CATEGORICAL AND LIMITED DEPENDENT VARIABLES (1997).
186 See SCOTT LONG, REGRESSION MODELS FOR CATEGORICAL AND LIMITED DEPENDENT VARIABLES (1997).
187 Our randomized trial already provides some degree of comfort in drawing causal inferences.
The regression models that we have run show the same pattern. A partial exception is a binomial logit regression, for which we converted the five-point clarity scale into a binary variable to indicate whether the judge’s response was either relatively low (1 or 2) or relatively high (3, 4, or 5). This dichotomy reflects the legal line between relatively unclear and clear statutes that is drawn by the lexical ordering instructions for the treatment group. In Trade Name, the treatment variable for exposure to a lower-tier source is again not statistically significant ($p = 0.87$). In Election Law, however, the treatment variable is only weakly significant ($p = 0.096$).189 These results reflect that much of the apparent movement in clarity responses in the treatment group is at clarity levels 3 and above, rather than across the critical legal line between clear and unclear statutes. The treatment and control group distributions of clarity responses are displayed in Figures 2 and 3.

- **Clarity-plus-judgment effects.** Our other hypotheses require us to combine statutory clarity responses with final judgments, which complicates the analysis.

To make progress on Hypothesis 2, regarding judgment-flipping without clarity responses changing, we can isolate the judges who assigned clarity levels of 3 and higher. This group chose a clarity level sufficiently high to rule out the lawful consideration of lower-tier sources; the treatment group did so explicitly in light of their instructions. We can then compare the percentage (equivalent to the mean) of judges in each group who entered a judgment for the plaintiff(s) rather than the defendant. Finding such an effect would indicate a rather serious lexical ordering failure among some judges, in which final judgments are flipped by exposure to lower-tier sources despite consistently high levels of reported statutory clarity.

In Trade Name, the percentage of judges who entered pro-plaintiff judgments at clarity levels 3 and higher in the control group (8%) and in the treatment group (14%) is not significantly different ($p = 0.46$, two-sided $t$-test)—again, despite the exposure to pro-plaintiff committee reports.190 But in Election Law, this difference between pro-plaintiff judgments at high clarity levels in the control group (62%) and the treatment group (85%) is statistically significant at conventional levels ($p = 0.03$, two-sided $t$-test)191—again, consistent with some judges moving toward the agency’s pro-plaintiff position.192 For reasons already noted, this test cannot readily distinguish between judgment-flipping without changes in clarity and more complex movements from multiple points along a continuum from high clarity pro-plaintiff judgments to high clarity pro-defendant judgments.193 But those more complex movements would be a different type of lexical ordering failure, not a success.

188 [NOTE TO READERS: We are in the process of think through models that include interaction terms involving the judges’ responses to the question about the usefulness of legislative history.]

189 This value falls to $p = 0.110$ when we exclude two final-judgment responses that were difficult to read.

190 In Trade Name, the percentage of judges assigning clarity levels of 3 or higher was very similar in the control (72%) and treatment (71%) groups.

191 This value falls to $p = 0.0524$ when we exclude two final-judgment responses that were difficult to read.

192 In Election Law, the percentage of judges assigning clarity levels of 3 or higher was more noticeably different in the control (63%) and treatment (78%) groups. This difference is some reason to suspect more than judgment-flipping taking place.

193 See supra note 182.
Finally, on the convenient use of clarity tests under Hypotheses 3 (attraction) and Hypothesis 4 (backlash), we can begin by examining raw percentages. Histograms that combine clarity responses with final judgments are displayed in Figures 4 and 5. These figures do not show especially clear patterns of attraction or backlash from exposure to lower-tier sources of the specific type predicted by Hypotheses 3 and 4. But some of the numbers are nonetheless intriguing. [NOTE TO READERS: We are in the process of formulating statistical tests for Hypotheses 3 and 4.]

In Trade Name, the numbers leave space for the attraction theory in Hypothesis 3 but also for other explanations. Pro-defendant judgments at clarity level 3 do look lower in the treatment group, while pro-plaintiff judgments at clarity level 2 do look higher. Perhaps some judges conveniently lowered their clarity assessments from 3 to 2 in order to make legal room for the pro-plaintiff legislative history. However, even if those movements are significant, the raw percentages also indicate a drop in pro-defendant judgments at clarity level 2, and at this low clarity level judges are legally free to consider the legislative history and move to a pro-plaintiff judgment at clarity level 2. This is an alternative and entirely lawful explanation for a rise in pro-plaintiff judgments at clarity level 2.

But such lawful movement still cannot explain the apparent 20% drop in pro-defendant judgments at clarity level 3. If real, what explains this drop?

Strikingly, the raw percentages also depict a rise of about 20% in pro-defendant judgments at clarity level 4 in the treatment group. If real, such movement is consistent with the backlash theory in Hypothesis 4, given the small number of judges who could have moved down from a pro-defendant judgment at clarity level 5. But the hypothesized backlash is supposed to come from pro-defendant judges who would have assigned a clarity level of 1 or 2 absent exposure to the legislative history. And although we have already noticed a possible drop in pro-defendant judgments at clarity level 2, we have also emphasized an apparent 20% drop in pro-defendant judgments at clarity level 3. Therefore, any real increase in pro-defendant judgments at clarity level 4 could be explained by a decrease in pro-defendant judgments at clarity level 3, not 2. So the raw percentages cannot confirm a convenient boost in clarity levels to meet the legal threshold at which legislative history must be ignored, which is what Hypothesis 4 predicts. Instead, some judges who would have been entitled to ignore the legislative history anyway might have boosted their clarity assessment still higher, thereby creating an extra “clarity cushion” against the legal line at which they would be required to consider legislative history.

In Election Law, the backlash theory in Hypothesis 4 has very little space to operate. The treatment group entered few pro-defendant judgments at clarity levels 3 or higher, hence not many judges can possibly be moving to the defendant’s side at high clarity levels. In contrast, some form of attraction to the agency’s position is likely occurring, but not necessarily the version predicted by Hypothesis 3. True, the treatment group seems to move away from pro-defendant judgments at clarity level 3, where reliance on the agency is forbidden: the percentage of judges in that slot of the histogram is about 20% of the control group and 0% of the treatment group. But we cannot see a complementary increase in pro-plaintiff judgments at clarity levels 1 or 2. Perhaps more than one movement is occurring and Hypothesis 3’s attraction theory fits one of them, but we cannot be sure with these numbers.

194 For example, (1) some judges in the treatment group could be responding to the agency’s position by moving from pro-defendant judgments at clarity level 3 into pro-plaintiff judgments at clarity level 2, while at the same
Actually, in Election Law the largest spike to be explained is for pro-plaintiff judgments at clarity level 4. Few judges entered pro-plaintiff judgments at clarity level 5, so the best explanation for this apparent increase must come from judges who were otherwise less supportive of the plaintiffs’ position. And a simple explanation, consistent with the raw percentages and with our test of judgment-flipping under Hypothesis 2, is that some judges who were exposed to the agency’s pro-plaintiff position switched from pro-defendant judgments at clarity level 3 all the way to pro-plaintiff judgments at clarity level 4. Once again, more complex movements might explain these patterns, but without eliminating the strong indications of lexical ordering failures of some kind among some judges in Election Law.

Recall that clarity level 4 is high enough to make the agency’s position legally irrelevant. The results in Election Law thus suggest that some judges were indeed attracted to the agency’s position, contrary to the lexical ordering instructions, but in a way that would prevent these judges from relying on the agency when explaining their final judgments. Judges entering pro-plaintiff judgment’s at this relatively high clarity level would position themselves to contend that the statute already quite clearly supports the plaintiffs and, therefore, the agency’s pro-plaintiff position either should not be addressed, or supports the final judgment in the alternative. Such judicial positioning could qualify as a “convenient” use of the clarity test in a sense, albeit not the sense captured by the particular attraction theory in Hypothesis 3.

---

195 See supra note 194. Some of the movement here might be coming from judges who would have entered pro-defendant judgments at clarity level 2, which is another form of attraction not covered by Hypothesis 3.
196 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).
197 Cf. Contract Freighters, Inc. v. Secretary of Transp., 260 F.3d 858, 862 n.5 (8th Cir. 2001) (indicating that the court would defer to the agency’s position even if the statute were unclear); United States v. 25 Cases, More or Less, of an Article of Device, 942 F.2d 1179, 1182 (7th Cir. 1991) (same); Bimini Superfast Operations LLC v. Winkowski, 994 F. Supp. 2d 106, 125 (D.D.C. 2014) (same).
### Table 1: Trade Name—Ordered Probit Models of Clarity Responses

<table>
<thead>
<tr>
<th>Independent Variables</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
<th>Model 5</th>
<th>Model 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>0.223</td>
<td>0.258</td>
<td>0.259</td>
<td>0.314</td>
<td>0.236</td>
<td>0.352</td>
</tr>
<tr>
<td></td>
<td>(0.215)</td>
<td>(0.227)</td>
<td>(0.227)</td>
<td>(0.233)</td>
<td>(0.218)</td>
<td>(0.233)</td>
</tr>
<tr>
<td>LH useful (1–7)</td>
<td>-0.0119</td>
<td>-0.0124</td>
<td>-0.00687</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0662)</td>
<td>(0.0667)</td>
<td>(0.0673)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Text vague (1–7)</td>
<td>0.00617</td>
<td>0.0447</td>
<td></td>
<td>0.0481</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.109)</td>
<td>(0.115)</td>
<td></td>
<td>(0.118)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Midwest phase</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.456</td>
<td>0.479</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.325)</td>
<td>(0.334)</td>
</tr>
<tr>
<td>Federal judge</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.353</td>
<td>0.327</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.425)</td>
<td>(0.426)</td>
</tr>
<tr>
<td>Military judge</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.226</td>
<td>-0.238</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.385)</td>
<td>(0.384)</td>
</tr>
<tr>
<td>LH useful high (6, 7)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.201</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(-0.0686</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.252)</td>
</tr>
<tr>
<td>LH useful low (1, 2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.0545</td>
<td>-0.142</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.273)</td>
</tr>
</tbody>
</table>

Observations: 101 | 95 | 95 | 95 | 101 | 96

### Table 2: Election Law—Ordered Probit Models of Clarity Responses

<table>
<thead>
<tr>
<th>Independent Variables</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
<th>Model 5</th>
<th>Model 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>0.579**</td>
<td>0.586***</td>
<td>0.583**</td>
<td>0.644***</td>
<td>0.545**</td>
<td>0.614***</td>
</tr>
<tr>
<td></td>
<td>(0.217)</td>
<td>(0.227)</td>
<td>(0.227)</td>
<td>(0.233)</td>
<td>(0.219)</td>
<td>(0.232)</td>
</tr>
<tr>
<td>LH useful (1–7)</td>
<td>-0.0510</td>
<td>-0.0486</td>
<td>-0.0391</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0645)</td>
<td>(0.0648)</td>
<td>(0.0653)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Text vague (1–7)</td>
<td>-0.0452</td>
<td>0.00180</td>
<td></td>
<td>-0.0107</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.107)</td>
<td>(0.112)</td>
<td></td>
<td>(0.114)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Midwest phase</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.162</td>
<td>0.104</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.319)</td>
<td>(0.326)</td>
</tr>
<tr>
<td>Federal judge</td>
<td>-0.0931</td>
<td>-0.0782</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.412)</td>
<td>(0.413)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Military judge</td>
<td>-0.597</td>
<td>-0.620</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.379)</td>
<td>(0.378)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LH useful high (6, 7)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.110</td>
<td>0.00118</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.249)</td>
<td>(0.265)</td>
</tr>
<tr>
<td>LH useful low (1, 2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.292</td>
<td>0.316</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.273)</td>
<td>(0.287)</td>
</tr>
</tbody>
</table>

Observations: 101 | 96 | 96 | 96 | 101 | 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, which was measured on a five-point scale.
**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
4. Discussion

In this experiment, judges showed mixed success in carrying out lexical ordering instructions. The results in Election Law show fairly consistent warning signs of lower-tier sources influencing clarity assessments or final judgments when only top-tier sources should have been considered. Even if some or many judges ignored the agency position when the instructions required it, the evidence indicates that not every judge did so. The results in Trade Name do not carry such consistent warnings. Perhaps some judges failed to ignore the legislative history when the instructions required it, but such failures are difficult to confirm.

Overall the results offer reason to suspect that judges sometimes, but only sometimes, succeed at lexically ordering sources in statutory cases. When judges succeed, the hard trade-offs discussed in Part II must be confronted. When judges fail, there are open questions about how exactly the failure influences decisions. One research goal for the future should be to identify the sizes and boundaries of the domains in which judges are likely or unlikely to carry out lexically ordered decision tasks, and to more precisely identify the effects of failure.

Our results do intimate a curious variety of judicial reactions to lower-tier sources. There are grounds to continue searching for the convenient use of vague clarity tests to make attractive sources legally relevant and to make unattractive sources irrelevant. But our results also hint that some judges may create “clarity cushions”: attractive lower-tier sources might be made irrelevant as a matter of law yet available as alternative grounds for judgment, and unattractive lower-tier sources might be characterized as obviously irrelevant as a matter of law. In these ways judges can avoid explicitly relying on or grappling with controversial or conflicting sources, while those sources nonetheless influence judges. But not every important potential influence of lower-tier sources can be traced, confirmed, or refuted here.

Our experiment has several limitations in addition to standard concerns about external validity. Although real cases often are more complex than these vignettes and many judges face heavy docket loads, we hope that judges treat actual cases with adequate care and in a way that might enhance their ability to lexically order sources. In addition, the judges in our experiment were, on the whole, less experienced in their current positions than the average judge. Perhaps more-experienced judges deal with lexical ordering instructions differently. If greater judicial experience increases the likelihood of successful lexical ordering, then we have another reason to continue thinking about the hard-trade-offs associated with this logical architecture. Although gathering judge participants is challenging, adding more federal judges and judges from other parts of the judicial hierarchy would be useful. Moreover, limited data on our respondent judges and limited variations in our vignettes leave open opportunities for future research.

Indeed our results are sufficiently intriguing to prompt thoughts about why judges might respond differently in these two case vignettes. First, as a category of sources, agency positions might be more attractive to more judges than committee reports. Additional experiments could test the thought by varying the type of lower-tier source across vignettes. Second, the content of

\[\text{198} \] We cannot be sure, at this stage, that failed lexical ordering equals unfettered aggregation of sources.

\[\text{199} \] The resulting opinions explaining final judgments presumably will be affected when clarity perceptions about top-tier sources are actually influenced by lower-tier sources. Apart from misdirecting the readers of those opinions regarding law’s clarity, the opinions could influence subsequent decisions. If taken seriously, these opinions might suggest that a certain number of superior sources with a certain modest level of implication in one direction should be considered more persuasive and clearer than the ideal assessment of those sources.
the agency position included a brief policy rationale, while the committee reports did not. Additional experiments could vary the content of various sources across case vignettes.\textsuperscript{200} Third, the arguable ideological charge of the issue in Election Law might matter, softening the force of the instructions yet encouraging reliance on a secondary source of law. This explanation is not purely attitudinal: If the judges’ decisions were driven solely by judicial ideology and situation sense, then adding an agency position might well make no difference. Either way, adding vignettes with varying ideological salience and additional extralegal facts would be useful.

Moreover, our control group was far more likely to be persuaded by the plaintiffs’ position in Election Law than in Trade Name. Perhaps in Election Law more judges were open to or on the verge of adopting the plaintiff’s position at baseline, and thus exposure to an agency position (and explanation) became too attractive to ignore. One might vary the direction of the lower-tier source content, pro-defendant and pro-plaintiff. In any event, some results might be a function of the distribution of responses in the control group. Note as well that the vignettes’ order of presentation was fixed and could be varied to study order effects. Perhaps judges lean toward judgment for a plaintiff after having entered judgment for a defendant, for example.

Finally, note that our experiment did not test the influence of legislative history or agency positions under an inclusive add-it-all-up \textit{American Trucking} approach. Instead our experiment compared the outright exclusion of those sources with a lexical ordering effort. In realistically evaluating lexical ordering, we should achieve a better sense of how judging would differ under all plausible alternatives for interpretive method. Perhaps an all-inclusive interpretive doctrine would yield patterns of results not so different from what we find under lexical ordering efforts, without the same chances of implicit attraction, backlash, and clarity cushions.

\textbf{D. Moving Forward}

[Omitted]

\textbf{CONCLUSION}

The defense of lexically ordered statutory interpretation today seems 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 can carry out these instructions; under other conditions, we should expect that the instructions will fail or make no difference at final judgment. When the instructions are followed and influential, the trade-offs often should be troubling, and yet that particular mix of trade-offs might well be avoided—either by simplifying 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, even at this late date, its popularity now exceeds our ability to confidently justify its continuation. The spread and persistence of lexical ordering in statutory interpretation should generate powerful demands for more knowledge, however acquired, about its real-world operation by judges.

\textsuperscript{200} The agency’s rationale, not necessarily the agency’s position itself, might explain its influence. If so, then a party’s articulation of the same rationale could have the same influence. Furthermore, the lexical ordering 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 in that event. Although the instructions roughly reflect contemporary doctrine, perhaps these formulations somehow influenced judges’ likelihood of ignoring the lower-tier sources in accord with the instructions.
APPENDIX—VIGNETTES

Case #1

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? . . .
Case #2

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? . . .