TESTING ORDINARY MEANING

Kevin P. Tobia

CONTENTS

TESTING ORDINARY MEANING

Kevin P. Tobia[*]

Within legal scholarship and practice, among the most pervasive tasks is the interpretation of texts. And within legal interpretation, perhaps the most pervasive inquiry is the search for "ordinary meaning." Jurists often treat ordinary meaning analysis as an empirical inquiry, aiming to discover a fact about how people understand language. When evaluating ordinary meaning, interpreters rely on dictionary definitions or patterns of common usage, increasingly via "legal corpus linguistics" approaches. However, the most central question about these popular methods remains open: Do they reliably reflect ordinary meaning? This Article presents experiments that assess whether (a) dictionary definitions and (b) common usage data reflect (c) how people actually understand language today.

The Article elaborates the implications of two main experimental results. First, neither the dictionary nor legal corpus linguistics methods reliably track ordinary people's judgments about meaning. This finding shifts the argumentative burden to jurists who rely on these tools to identify "ordinary meaning" or "original public meaning": these views must articulate and demonstrate a reliable method of analysis. Moreover, this divergence illuminates several interpretive fallacies. For example, advocates of legal corpus linguistics often contend that the nonappearance of a specific use in a corpus indicates that the use is not part of the relevant term's ordinary meaning. The experiments reveal this claim to be a "Nonappearance Fallacy." Ordinary meaning exceeds datasets of common usage — even very large ones.

Second, dictionary and legal corpus linguistics verdicts diverge dramatically from each other. Part of that divergence is explained by the finding that broad dictionary definitions tend to direct interpreters to extensive interpretations, while data of common usage tends to point interpreters to more prototypical cases. This divergence suggests two different criteria that are often relevant in interpretation: a more extensive criterion and a more narrow criterion. Although dictionaries and legal corpus linguistics might, in some cases, help us identify these criteria, a hard legal-philosophical question remains: Which of these two criteria should guide the interpretation of terms and phrases in legal texts? Insofar as there is no compelling case to prefer one, the results suggest that dictionary definitions, legal corpus linguistics, or even other more scientific measures of meaning may not be equipped in principle to deliver simple and unequivocal answers to inquiries about the so-called "ordinary meaning" of legal texts.

[*] Assistant Professor of Law, Georgetown University Law Center; Research Affiliate, ETH Zurich Center for Law & Economics. Many thanks to the editors of the *Harvard Law Review*; to audiences at the Association for American Law Schools, Experimental Methods in Legal Scholarship, American Law and Economics Association, and Law & Corpus Linguistics at BYU conferences; and for thoughtful comments from Ellen Aprill, Elliott Ash, Stefan Bechtold, Clark Cunningham, Jesse Egbert, Richard Elliott, William Eskridge, Ed Finnegan, Tammy Gales, Stephen Galoob, Tobias Gesche, Abbe Gluck, Neal Goldfarb, Stefan Gries, Ivar Hannikainen, David Hoffman, Daniel Keller, Josh Kleinfeld, Markus Kneer, Josh Knobe, Issa Kohler-Hausmann, Anita Krishnakumar, Thomas Lee, Tom Lyon, Jamie Macleod, Stefan Magen, Daniel Markovits, John Mikhail, Stephen Mouritsen, Victoria Nourse, Karolina Prochownik, Kevin Reuter, Martin Schonger, Scott Shapiro, Dan Simon, Brian Slocum, Larry Solan, Roseanna Sommers, Alexander Stremitzer, Mila Versteeg, Pascale Willemsen, Gideon Yaffe, and Evan Zoldan.

INTRODUCTION

Within legal scholarship and practice, among the most pervasive tasks is the interpretation of texts. And within legal interpretation, among the most pervasive inquiries is the search for ordinary meaning.[1] Across the interpretation of contracts,[2] wills,[3] trusts,[4] deeds,[5] patents,[6] statutes,[7] regulations,[8] treaties,[9] and constitutions,[10] legal theorists and practitioners regularly evaluate the text's ordinary meaning.

This Article focuses primarily on interpretation of American contracts and statutes, but ordinary meaning is also of global legal significance: "[E]very legal system recognizes the importance of ordinary meaning . . . ."[11] This is for good reason: "[W]hat method of . . . interpretation

---

[1] *See generally* BRIAN G. SLOCUM, ORDINARY MEANING: A THEORY OF THE MOST FUNDAMENTAL PRINCIPLE OF LEGAL INTERPRETATION (2015).

[2] *See, e.g.*, CAL. CIV. CODE § 1644 (West 2018) ("The words of a contract are to be understood in their ordinary and popular sense . . . ."); Jowett, Inc. v. United States, 234 F.3d 1365, 1369 (Fed. Cir. 2000) ("We give the words of the agreement their ordinary meaning unless the parties mutually intended and agreed to an alternative meaning." (quoting Harris v. Dep't of Veterans Affs., 142 F.3d 1463, 1467 (Fed. Cir. 1988))).

[3] *See, e.g.*, CAL. PROB. CODE § 21122 (West 2018) ("The words of an instrument are to be given their ordinary and grammatical meaning unless the intention to use them in another sense is clear and their intended meaning can be ascertained.").

[4] *See, e.g.*, *id.*

[5] *See, e.g.*, Lambert v. Pritchett, 284 S.W.2d 90, 90 (Ky. 1955) ("Terms are to be construed and understood according to their plain, ordinary, and popular sense . . . ."); Burdette v. Bruen, 191 S.E. 360, 363 (W. Va. 1937).

[6] *See, e.g.*, Phillips v. AWH Corp., 415 F.3d 1303, 1321 (Fed. Cir. 2005) ("[T]he 'ordinary meaning' of a claim term is its meaning to the ordinary artisan . . . .").

[7] *See, e.g.*, Moskal v. United States, 498 U.S. 103, 108 (1990) ("'In determining the scope of a statute, we look first to its language,' giving the 'words used' their 'ordinary meaning.'" (citations omitted) (first quoting United States v. Turkette, 452 U.S. 576, 580 (1981); and then twice quoting Richards v. United States, 369 U.S. 1, 9 (1962))).

[8] *See, e.g.*, Bowles v. Seminole Rock & Sand Co., 325 U.S. 410, 414 (1945) (noting that the Court's only tools "are the plain words of the regulation and any relevant interpretations of the Administrator"). *See generally* Christensen v. Harris County, 529 U.S. 576 (2000).

[9] *See, e.g.*, Vienna Convention on the Law of Treaties art. 31, *opened for signature* May 23, 1969, 1155 U.N.T.S. 331, 340 ("A treaty shall be interpreted in good faith in accordance with the ordinary meaning to be given to the terms of the treaty in their context and in light of its object and purpose."). *See generally* Curtis J. Mahoney, *Treaties as Contracts: Textualism, Contract Theory, and the Interpretation of Treaties*, 116 YALE L.J. 824 (2007); Brian G. Slocum & Jarrod Wong, *The Vienna Convention and the Ordinary Meaning of International Law*, 46 YALE J. INT'L L. (forthcoming 2021).

[10] *See generally, e.g.*, Lawrence B. Solum, The Constraint Principle: Original Meaning and Constitutional Practice (Apr. 3, 2019) (unpublished manuscript) (on file with the Harvard Law School Library).

[11] Daniel A. Farber, *The Hermeneutic Tourist: Statutory Interpretation in Comparative Perspective*, 81 CORNELL L. REV. 513, 516 (1996) (reviewing INTERPRETING STATUTES: A COMPARATIVE STUDY (D. Neil MacCormick & Robert S. Summers eds., 1991)) (discussing a study that examined the relevance of ordinary meaning to legal interpretation in jurisdictions including Argentina, Britain, Finland, France, Germany, Italy, Poland, Sweden, and the United States).

would view the ordinary meaning of words as completely irrelevant?"[12] Even legal theorists who advocate looking beyond ordinary meaning acknowledge that, in interpretation, "one certainly begins there."[13]

Despite this general agreement concerning ordinary meaning's legal relevance, there is significant debate about how exactly to elaborate the concept of ordinary meaning. On one common view, the ordinary meaning of a text is what its words would communicate to ordinary people. In some circumstances, legal theories seek the *original* ordinary meaning or "original public meaning" of a text: what its words would have communicated to people at some past time, such as the time a contract or will is formalized, a bill becomes a statute, or a constitution or treaty is ratified.[14] Various legal debates concern this original or historical ordinary meaning of a text, especially in statutory and constitutional contexts,[15] but also in contractual ones.[16]

Analysis of ordinary meaning is extraordinarily common — and increasingly so. The ubiquity of ordinary meaning analysis can be explained in part by the ubiquity of legal interpretation,[17] where ordinary meaning is "the most fundamental principle."[18] Analysis of the ordinary, plain, or natural meaning underlies popular approaches to

---

[12] *Id.* (citing Edward Rubin, Book Review, 41 AM. J. COMPAR. L. 128, 139 (1993) (reviewing INTERPRETING STATUTES, *supra* note 11)).

[13] Felix Frankfurter, *Some Reflections on the Reading of Statutes*, 47 COLUM. L. REV. 527, 535 (1947).

[14] *See* Thomas R. Lee & Stephen C. Mouritsen, *Judging Ordinary Meaning*, 127 YALE L.J. 788, 826 (2018).

[15] Debate is especially divisive in the statutory and constitutional context. *See generally, e.g.*, Victoria Nourse, *Reclaiming the Constitutional Text from Originalism: The Case of Executive Power*, 106 CALIF. L. REV. 1 (2018); Solum, *supra* note 10. Following Justice Lee and Professor Stephen Mouritsen, I use "original public meaning" to refer to a legal text's communicative content (or "ordinary meaning") at the relevant time. *See* Lee & Mouritsen, *supra* note 14, at 825–26. For example, the original public meaning of a 1967 statute is that text's ordinary meaning in 1967.

[16] *See, e.g.*, Stephen C. Mouritsen, *Contract Interpretation with Corpus Linguistics*, 94 WASH. L. REV. 1337 (2019).

[17] *See generally, e.g.*, AHARON BARAK, PURPOSIVE INTERPRETATION IN LAW (2005); WILLIAM N. ESKRIDGE, JR., INTERPRETING LAW: A PRIMER ON HOW TO READ STATUTES AND THE CONSTITUTION (2016); ROBERT A. KATZMANN, JUDGING STATUTES (2014); ANTONIN SCALIA & BRYAN A. GARNER, READING LAW: THE INTERPRETATION OF LEGAL TEXTS (2012); SLOCUM, *supra* note 1; LAWRENCE M. SOLAN, THE LANGUAGE OF JUDGES (1993); LAWRENCE M. SOLAN, THE LANGUAGE OF STATUTES: LAWS AND THEIR INTERPRETATION (2010); Richard H. Fallon, Jr., *The Meaning of Legal "Meaning" and Its Implications for Theories of Legal Interpretation*, 82 U. CHI. L. REV. 1235 (2015).

[18] SLOCUM, *supra* note 1.

contract interpretation[19] in the United States and abroad,[20] which remains the "most important source of commercial litigation."[21] Ordinary meaning analysis also informs interpretation of patents,[22] trusts, and wills.[23]

An empirical study of the 2005–2008 Supreme Court Terms found that the majority of Supreme Court Justices "referenced text/plain meaning and Supreme Court precedent more frequently than any of the other interpretive tools."[24] In light of the three most recent Supreme Court appointments, ordinary meaning analysis will likely continue to hold a significant place.[25] Consider Justice Kavanaugh's view of interpreting statutes: "Under the 'best reading' inquiry, the question is only how the words [of a statute] would be read by an ordinary user of the English language. That's why textualists rely on dictionaries."[26] Similarly, Justice Gorsuch frequently assesses the ordinary meaning of legal texts. In his first Supreme Court opinion, he analyzed a statute's "ordinary meaning," citing both (i) the Oxford English Dictionary as well as (ii) common patterns of language use.[27]

Ordinary meaning analysis is sometimes associated with conservative legal thought. But the consideration of ordinary meaning is broader. In *Muscarello v. United States*,[28] Justice Breyer analyzed the statute's "ordinary" meaning, employing similar methods to those utilized by Justice Gorsuch.[29] Justice Breyer referred to (i) dictionary

---

19 *See generally, e.g.*, Shawn Bayern, *Contract Meta-interpretation*, 49 U.C. DAVIS L. REV. 1097, 1099 (2016); Ronald J. Gilson et al., *Contract and Innovation: The Limited Role of Generalist Courts in the Evolution of Novel Contractual Forms*, 88 N.Y.U. L. REV. 170, 171 n.1 (2013) ("Textualist interpretation . . . looks to a contract's formal language . . . ."); Robert E. Scott, *The Death of Contract Law*, 54 U. TORONTO L.J. 369 (2004); Gregory Klass, Contract Exposition and Formalism (2017) (unpublished manuscript) (on file with the Harvard Law School Library).

20 *See, e.g.*, Arnold v. Britton [2015] UKSC 36 (appeal taken from EWCA (Civ.)) ("[T]he reliance placed in some cases on commercial common sense and surrounding circumstances . . . should not be invoked to undervalue the importance of the language of the provision which is to be construed.").

21 Ronald J. Gilson, Charles F. Sabel & Robert E. Scott, *Text* and *Context: Contract Interpretation as Contract Design*, 100 CORNELL L. REV. 23, 25 (2014).

22 *See generally, e.g.*, Tun-Jen Chiang & Lawrence B. Solum, *The Interpretation-Construction Distinction in Patent Law*, 123 YALE L.J. 530 (2013).

23 *E.g.*, CAL. PROB. CODE § 21122 (West 2018); NAOMI R. CAHN, ALYSSA A. DIRUSSO & SUSAN N. GARY, WILLS, TRUSTS, AND ESTATES IN FOCUS 144 (2019).

24 Anita S. Krishnakumar, *Statutory Interpretation in the Roberts Court's First Era: An Empirical and Doctrinal Analysis*, 62 HASTINGS L.J. 221, 251 (2010).

25 *See, e.g.*, Brett M. Kavanaugh, *Fixing Statutory Interpretation*, 129 HARV. L. REV. 2118, 2125 (2016) (reviewing ROBERT A. KATZMANN, JUDGING STATUTES (2014)).

26 *Id.* at 2150 n.158 ("Dictionaries may not provide authoritative, binding interpretations of the language of a statute, but they do tell courts something about how the ordinary user of the English language might understand that statutory language.").

27 Henson v. Santander Consumer USA Inc., 137 S. Ct. 1718, 1722–24 (2017).

28 524 U.S. 125 (1998).

29 *Id.* at 127–28.

definitions[30] and (ii) patterns of common word usage in "computerized newspaper databases."[31]

Similarly, although ordinary meaning analysis is often associated with textualism and formalism,[32] a diverse range of theories endorse the relevance of ordinary meaning as one criterion in legal interpretation.[33] And on other views, ordinary meaning is crucial evidence of other criteria: when interpreting contracts by aiming to uncover and preserve the parties' *mutual intent*, "the words of an integrated agreement remain the most important evidence of intention."[34]

Of course, emphasis on ordinary meaning is not central to all theories of interpretation, and notable detractors question the empirical assumptions required to discover ordinary meaning: Do judges actually have the ability, insight, or tools to determine the ordinary meaning of legal texts?[35]

That critique highlights a crucial insight. Ordinary meaning inquiries are often understood as *empirical* ones, which aim to discover descriptive facts about meaning.[36] Theories holding that a legal text must be applied consistently with its ordinary meaning do not typically characterize their project as a normative inquiry. Rather than debating how a text *should* be understood by some ideal person, these theories ask how a text would *in fact* be understood by ordinary people.

There are several empirical methods commonly used to inquire into a text's ordinary meaning, including consulting dictionary definitions or using "legal corpus linguistics" to analyze patterns of language usage

---

[30] *Id.* at 128–29.

[31] *Id.* at 129.

[32] *See* Victoria Nourse, *Textualism 3.0: Statutory Interpretation After Justice Scalia*, 70 ALA. L. REV. 667, 681 (2019). *See generally* Klass, *supra* note 19 (discussing interpretive formalism in contract law).

[33] *See, e.g.*, ESKRIDGE, *supra* note 17, at 35 ("There are excellent reasons for the primacy of the ordinary meaning rule."); Lee & Mouritsen, *supra* note 14, at 788 (claiming that "most everyone agrees" that the ordinary meaning rule predominates in legal interpretation); Lawrence B. Solum, *Surprising Originalism: The Regula Lecture*, 9 CONLAWNOW 235, 251–59 (2018); *see also* Harvard Law School, *The Scalia Lecture: A Dialogue with Justice Kagan on the Reading of Statutes*, YOUTUBE, at 8:29 (Nov. 25, 2015), https://youtu.be/dpEtszFToTg [http://perma.cc/3BCF-FEFR] ("We're all textualists now.").

[34] RESTATEMENT (SECOND) OF CONTRACTS § 212 cmt. b (AM. L. INST. 1981). And in determining the *purpose* of a statute, purposivists often ask, "what would a reasonable human being intend this specific language to accomplish?" Stephen Breyer, Lecture, *On the Uses of Legislative History in Interpreting Statutes*, 65 S. CAL. L. REV. 845, 854 (1992) (internal quotations omitted). *See generally* HENRY M. HART, JR. & ALBERT M. SACKS, THE LEGAL PROCESS: BASIC PROBLEMS (1995).

[35] *See, e.g.*, Fallon, *supra* note 17, at 1235; Cass R. Sunstein, *There is Nothing that Interpretation Just Is*, 30 CONST. COMMENT. 193, 194–95 (2015).

[36] *See, e.g.*, Randy E. Barnett, *Interpretation and Construction*, 34 HARV. J.L. & PUB. POL'Y 65, 66 (2011) ("It cannot be overstressed that the activity of determining semantic meaning at the time of enactment required by the first proposition is *empirical*, not normative." (citing KEITH E. WHITTINGTON, CONSTITUTIONAL INTERPRETATION: TEXTUAL MEANING, ORIGINAL INTENT, AND JUDICIAL REVIEW 6 (1999))).

across a corpus.[37] The popularity of these methods is not difficult to explain. Dictionary use and the dominant form of legal corpus linguistics are both relatively easy to employ. Moreover, they often seem objective, neutral, and scientific.[38]

Both methods are also increasingly popular. The Supreme Court cites dictionaries more today than ever before.[39] Legal corpus linguistics is certainly less prevalent, but it has also grown in use and esteem.[40] The Supreme Court has examined patterns of word use through newspaper databases,[41] and state supreme courts have searched corpora including the Corpus of Contemporary American English (COCA).[42] The

---

[37] There is a growing literature on legal corpus linguistics. This Article is primarily concerned with what is currently the most popular and prevalent form of legal corpus linguistics, which seeks to identify "ordinary meaning" via evidence of common usage (for example, data about the frequency of word usage). This approach is exemplified by Justice Lee and Professor Mouritsen's work. *See* Lee & Mouritsen, *supra* note 14. This is also the form of corpus linguistics that has been used or cited thus far in actual legal decisions. *See, e.g.*, Carpenter v. United States, 138 S. Ct. 2206, 2238–39 (2018) (Thomas, J., dissenting) (citing corpus linguistic evidence that "[t]he phrase 'expectation(s) of privacy' does not appear in" Founding-era sources, *id.* at 2238); Caesars Ent. Corp. v. Int'l Union of Operating Eng'rs, 932 F.3d 91, 95 (3d Cir. 2019) (citing "most common synonyms" of the relevant term, "previously," and the "words that most often co-occurred" with it); Wilson v. Safelite Grp., Inc., 930 F.3d 429, 444 (6th Cir. 2019) (Thapar, J., concurring in part and concurring in the judgment) (taking frequency data as evidence of ordinary meaning); State v. Lantis, 447 P.3d 875, 880–81 (Idaho 2019) (noting that in the corpus linguistics search concerning the phrase "disturbing the peace," "88.4% referenced a public, external, physical peace," *id.* at 881, and that this finding supported the court's "conclusion that 'disturbing the peace' has a meaning that nearly always refers to public, external peace," *id.*); Richards v. Cox, 450 P.3d 1074, 1079 (Utah 2019) (citing frequency data to interpret the meaning of "employment"); Fire Ins. Exch. v. Oltmanns, 416 P.3d 1148, 1163 n.9 (Utah 2018) (Durham, J., concurring in part and concurring in the result) (advocating corpus linguistics as a tool to identify the "most frequent meaning" and "most common meaning" (internal quotations omitted)); State v. Rasabout, 356 P.3d 1258, 1275–82 (Utah 2015) (Lee, A.C.J., concurring in part and concurring in the judgment) (taking frequency data as evidence of ordinary meaning); *In re* Adoption of Baby E.Z., 266 P.3d 702, 725–26 (Utah 2011) (Lee, J., concurring in part and concurring in the judgment) (same).

This Article refers to this particular form of corpus linguistics as "legal corpus linguistics." It is, of course, crucial to acknowledge the possibility that some new method of legal corpus linguistics might be developed. Moreover, such a method might, perhaps, be shown to escape the problems raised here. The Article's burden-shifting conclusion, *see infra* section VI.C, pp. 798–801, invites precisely this response. The Article's critiques should therefore not be seen as critiques of the broader linguistic field of "corpus linguistics," which contributes to many other questions beyond legal debates about "ordinary meaning" and "original public meaning."

[38] *See, e.g.*, Lawrence M. Solan, *Can Corpus Linguistics Help Make Originalism Scientific?*, 126 YALE L.J.F. 57 (2016).

[39] *See* John Calhoun, Note, *Measuring the Fortress: Explaining Trends in Supreme Court and Circuit Court Dictionary Use*, 124 YALE L.J. 484, 497, 497 fig.1, 502 fig.3 (2014).

[40] *See generally* Lee & Mouritsen, *supra* note 14; Lee J. Strang, *How Big Data Can Increase Originalism's Methodological Rigor: Using Corpus Linguistics to Reveal Original Language Conventions*, 50 U.C. DAVIS L. REV. 1181 (2017); Evan C. Zoldan, *Corpus Linguistics and the Dream of Objectivity*, 50 SETON HALL L. REV. 401 (2019).

[41] Muscarello v. United States, 524 U.S. 125, 129 (1998).

[42] *See, e.g.*, People v. Harris, 885 N.W.2d 832, 839 (Mich. 2016).

growing use of dictionaries and legal corpus linguistics is likely to continue.[43]

Yet, despite the enthusiasm surrounding dictionaries and legal corpus linguistics, there is surprisingly little work assessing what these tools actually do in legal interpretation. Although the use of dictionaries and legal corpus linguistics seems to grow more sophisticated,[44] their reliability has never been rigorously assessed.[45] There are important critiques of these methods from external theoretical perspectives,[46] but we might also take an internal perspective, considering whether these methods succeed on their own terms. Theories relying on these tools typically *assume* that dictionaries and legal corpus linguistics reliably reflect ordinary meaning,[47] what "the ordinary user of the English language might understand,"[48] but the question remains: Is this assumption true?

---

[43] Just last year, corpus linguistic analysis appeared twice in opinions from Judge Thapar of the Sixth Circuit. Judge Thapar issued a concurring opinion that relied on his corpus linguistics analysis, *see* Wilson v. Safelite Grp., Inc., 930 F.3d 429, 444 (6th Cir. 2019) (Thapar, J., concurring in part and concurring in the judgment), and a recent order requested that parties provide supplemental briefing that includes an explanation of how the corpus of Founding-era American English bears on the questions presented, *see* Wright v. Spaulding, 939 F.3d 695, 700 n.1 (6th Cir. 2019).

The Supreme Court's 2019 Term included a number of blockbuster interpretation cases, including whether the Eighth and Fourteenth Amendments permit states to abolish the insanity defense, Kahler v. Kansas, 140 S. Ct. 1021 (2020), and whether Title VII's prohibition of discrimination "because of . . . sex," 42 U.S.C. § 2000e-2(a)(1), prohibits discrimination against employees for being lesbian, gay, or transgender, Bostock v. Clayton County, 140 S. Ct. 1371 (2020). Corpus linguistics scholars filed a brief in the latter case. *See* Brief for Corpus-Linguistics Scholars Professors Brian Slocum, Stefan Th. Gries, and Lawrence Solan as Amici Curiae in Support of Employees, Bostock v. Georgia, 139 S. Ct. 1599 (2019) (No. 17-1618).

[44] *See, e.g.*, Lee & Mouritsen, *supra* note 14, at 830–36; Lawrence M. Solan & Tammy Gales, *Corpus Linguistics as a Tool in Legal Interpretation*, 2017 BYU L. REV. 1311, 1337–41.

[45] By comparison, other approaches to judicial interpretation have been addressed by empirical studies. *See generally* Abbe R. Gluck & Lisa Schultz Bressman, *Statutory Interpretation from the Inside — An Empirical Study of Congressional Drafting, Delegation, and the Canons: Part I*, 65 STAN. L. REV. 901 (2013) [hereinafter Gluck & Bressman, *Part I*]; Lisa Schultz Bressman & Abbe R. Gluck, *Statutory Interpretation from the Inside — An Empirical Study of Congressional Drafting, Delegation, and the Canons: Part II*, 66 STAN. L. REV. 725 (2014). There are two published experimental surveys about originalism. *See* Donald L. Drakeman, *What's the Point of Originalism?*, 37 HARV. J.L. & PUB. POL'Y 1123 (2014); Jamal Greene, Nathaniel Persily & Stephen Ansolabehere, *Profiling Originalism*, 111 COLUM. L. REV. 356 (2011). These are fascinating studies, but neither tests the reliability of originalist/textualist methodology. Instead, they focus on questions such as why people are originalists. *See, e.g.*, Greene et al., *supra*, at 359.

[46] *See generally* Ethan J. Herenstein, *The Faulty Frequency Hypothesis: Difficulties in Operationalizing Ordinary Meaning Through Corpus Linguistics*, 70 STAN. L. REV. ONLINE 112 (2017); Carissa Byrne Hessick, *Corpus Linguistics and the Criminal Law*, 2017 BYU L. REV. 1503; Victoria Nourse, *Picking and Choosing Text: Lessons for Statutory Interpretation from the Philosophy of Language*, 69 FLA. L. REV. 1409 (2017); Brian G. Slocum, *Ordinary Meaning and Empiricism*, 40 STATUTE L. REV. 13 (2019).

[47] For a demonstration of the use of corpus linguistics, see Lee & Mouritsen, *supra* note 14, at 836–45. However, that demonstration is not necessarily a demonstration of a *reliable* method of corpus linguistics.

[48] Kavanaugh, *supra* note 25, at 2150 n.158.

This Article develops a novel method to test dictionaries and legal corpus linguistics. This method provides evidence about the reliability of these tools and, by extension, of the theories that rely on them. Insofar as a legal interpretive theory relies upon dictionary definitions or patterns of word usage, the study here also provides evidence about the success of such legal theories. Part I outlines the background to these debates and legal theories in which ordinary meaning analysis is significant. Part II surveys the tools that provide interpretive evidence for those theories, including the consultation of dictionary definitions and patterns of linguistic usage across corpora.

Parts III and IV consider the accuracy and reliability of these putative measures of ordinary meaning. One reason that these tools have not yet been tested is that such a test may seem impossible, particularly when we are considering *original* meaning in some historical time period. How can we evaluate the accuracy of an eighteenth-century dictionary or even a corpus linguistics search from the 1980s without facts about the way in which the contested term was actually understood at the time (that is, without the best data about its original ordinary meaning)? However, what we can do is evaluate whether *modern* uses of contemporary dictionaries and legal corpus linguistics reflect terms' *modern* meanings. If a tool (for example, dictionary use) performs poorly in modern interpretation, so long as there are no historically distinguishing factors (for example, reasons that use of an eighteenth-century dictionary would be *more* reliable in historical interpretation than use of a modern dictionary is in modern interpretation), this fact gives us some evidence that the method is also unreliable in estimating original meaning.

To answer this central question — do dictionaries and legal corpus linguistics reliably reflect ordinary meaning? — the Article presents a robust experimental investigation of dictionaries and legal corpus linguistics. Experimental studies of ordinary people, law students, and United States judges all indicate that the way people understand ordinary terms and phrases (for example, "vehicle" or "carrying a firearm") varies systematically from what a dictionary definition or relevant legal corpus linguistics' usage data would indicate about the meaning.

After presenting and analyzing a range of experiments, the Article identifies ten noteworthy results and elaborates on their implications. Judges, law students, and ordinary people were strikingly similar in their ordinary conceptual judgments, use of dictionaries, and use of legal corpus linguistics. But *within* each group there existed considerable disagreement. For example, subjects within each group were divided on whether a canoe is a vehicle.

Legal corpus linguistics users tended to identify narrow, prototypical examples (for example, a car is a vehicle) rather than nonprototypical examples (for example, an airplane is a vehicle), while dictionary users tended in the opposite direction, supporting more extensive judgments

(for example, a pair of roller skates is a vehicle). As such, the verdicts of legal corpus linguistics and dictionary use often diverged from each other.

Moreover, both legal corpus linguistics and dictionary use diverged from ordinary meaning — at a conservative divergence rate between 20–35%. That is, across all the levels of expertise — ordinary people, law students, and judges — the data suggest that relying on *just* a dictionary definition or selection of legal corpus linguistics data would lead users to the wrong judgment about "ordinary meaning" fairly often, once in every three to five cases. And that conservative rate may not tell the whole story. The experiments included a number of relatively easy categorizations, such as whether a car is a vehicle. Insofar as real legal decisions often concern comparatively harder categorizations, such as whether an airplane is a vehicle, it is also instructive to consider the maximum error rate: For example, what percentage of judges using dictionary or legal corpus linguistics methods evaluated *the hardest interpretive question* incorrectly? Across all levels of expertise, relying on a dictionary definition or corpus linguistics data led 80–100% of users to the incorrect verdict.

For theories committed to the notion of a single "ordinary meaning" that determines legal outcomes across a range of cases and contexts, the results suggest that dictionaries and legal corpus linguistics — two central tools of discovering ordinary meaning — are unreliable in interpretation. This shifts the argumentative burden to those who rely on these tools to provide both a principled defense of their use *and* an empirical demonstration of how error can be avoided. Without such an account and demonstration, it is hard to take seriously claims that these tools should be used by judges to determine "ordinary meaning" or that these tools deliver objective, restrained, or nondiscretionary interpretive outcomes.

Moreover, the results help identify common fallacies in the use of dictionaries and legal corpus linguistics. For instance, consider what this Article calls the "Nonappearance Fallacy" — namely, the (false) claim that absence of a usage from a large corpus indicates that the usage is not part of the ordinary meaning. One might be surprised to learn that in some modern corpora, there are *no* airplanes referred to as "vehicles." Although it is tempting to conclude from this that airplanes are not included in the ordinary meaning of "vehicles," the experiments show this to be too quick: the majority of ordinary people, law students, and judges evaluated airplanes to be vehicles. The broader insight underlying these critiques is that *ordinary meaning diverges from ordinary use*. Although courts and commentators sometimes conflate these concepts, there is a crucial distinction between ordinary meaning and what is typically spoken and recorded (for example, in a corpus).

For theories uncommitted to, or even skeptical of, the notion of a single "ordinary meaning," the results illuminate two different criteria

that are often relevant in assessments of the meaning of legal texts: a more extensive criterion and a more narrow, prototypical criterion. For example, an extensive criterion indicates that airplanes, canoes, and even drones are vehicles, while a more prototypical criterion indicates that these entities are not vehicles (only cars and similar entities are vehicles). The findings show that dictionaries or legal corpus linguistics sometimes help clarify *one* of these criteria — often dictionaries point to the extensive criterion and legal corpus linguistics to the prototypical one — but a hard legal-philosophical question remains: Which of these *should* serve as a criterion in legal interpretation? Good reasons underlie both criteria, especially the fact that ordinary people may understand the relevant concepts with *both*. As such, the results suggest that dictionary definitions, legal corpus linguistics, or even other more scientific measures of meaning may not be equipped *in principle* to deliver simple and unequivocal answers to inquiries about the ordinary meaning of law. Instead, in hard cases, legal interpreters will likely have to look beyond the dictionary and corpus — to the legal text's context, history, and purpose.

I. ORDINARY MEANING

A. The Concept of Ordinary Meaning

What is ordinary meaning? As many have noted, terms like "ordinary meaning," "(original) public meaning," and "plain meaning" can be unclear.[49] Many treat "ordinary" and "plain" meaning synonymously.[50] And recent work on legal corpus linguistics in originalist and textualist interpretation treats "ordinary" and "public" meaning coextensively when considering nonspecialized terms and phrases.[51]

Although there may be important differences among these various concepts, it is also worth reflecting upon what unifies them. Each aims to capture the legal significance of *ordinary people's understanding* of legal texts. The wide-reaching legal significance of ordinary understanding is well put by Justice Holmes:

> [In contract interpretation] we ask, not what this man meant, but what those words would mean in the mouth of a normal speaker of English . . . . In the case of a statute, to turn from contracts to the opposite extreme, it would be

---

[49] *E.g.*, WILLIAM N. ESKRIDGE, JR., ET AL., CASES AND MATERIALS ON LEGISLATION: STATUTES AND THE CREATION OF PUBLIC POLICY 792–93 (4th ed. 2007); Lee & Mouritsen, *supra* note 14, at 798; Richard A. Posner, *Statutory Interpretation — In the Classroom and in the Courtroom*, 50 U. CHI. L. REV. 800, 808 (1983); *see also* SLOCUM, *supra* note 1; Fallon, *supra* note 17.

[50] *See, e.g.*, Mouritsen, *supra* note 16. Others draw a distinction such that "ordinary meaning" refers to a more decontextualized meaning while "plain meaning" refers to the meaning the relevant term, phrase, or clause has *within* the contract, statute, or other legal document.

[51] *See* Lee & Mouritsen, *supra* note 14.

> possible to say that as we are dealing with the commands of the sovereign the only thing to do is find out what the sovereign wants. . . . Yet in fact we do not deal differently with a statute from our way of dealing with a contract. We do not inquire what the legislature meant; we ask only what the statute means. . . . So in the case of a will. It is true that the testator is a despot, within limits, over his property, but he is required by statute to express his commands in writing, and that means that his words must be sufficient for the purpose when taken in the sense in which they would be used by the normal speaker of English under his circumstances.[52]

This Article does not take a position regarding whether ordinary meaning *should* serve as a criterion of legal interpretation in any domain. Rather, it develops implications for various competing theories of legal interpretation.[53] Nevertheless, there are various plausible reasons to support ordinary meaning analysis. For one, ordinary meaning is important because case law and "binding authoritative texts direct courts to consider it."[54] Legal interpretation that reflects ordinary meaning also promotes reliance values, fostering coordination and helping ordinary people manage their expectations.[55] Moreover, interpreting a text in line with its ordinary meaning may prevent adjudicators from imposing their own personal beliefs in interpretation[56] and limit opportunistic behavior by the document's drafters.[57] Empirical evidence suggests contracting parties might generally prefer contracts to be interpreted by their ordinary meanings.[58] Further, interpreting laws in accord with their ordinary meaning promotes the publicity and clarity of law.[59]

---

[52] Oliver Wendell Holmes, *The Theory of Legal Interpretation*, 12 HARV. L. REV. 417, 417, 419–20 (1899).

[53] *See infra* Part VI, pp. 789–804.

[54] James A. Macleod, *Ordinary Causation: A Study in Experimental Statutory Interpretation*, 94 IND. L.J. 957, 979 (2019).

[55] *See* Gilson et al., *supra* note 21, at 40–41; Scott, *supra* note 19, at 376.

[56] Uri Benoliel, *The Interpretation of Commercial Contracts: An Empirical Study*, 69 ALA. L. REV. 469, 475 (2017).

[57] *See* CATHERINE MITCHELL, INTERPRETATION OF CONTRACTS 113 (2007).

[58] *See* Benoliel, *supra* note 56, at 491–92; Lisa Bernstein, *Private Commercial Law in the Cotton Industry: Creating Cooperation Through Rules, Norms, and Institutions*, 99 MICH. L. REV. 1724, 1725, 1735 (2001); Lisa Bernstein, *Merchant Law in a Merchant Court: Rethinking the Code's Search for Immanent Business Norms*, 144 U. PA. L. REV. 1765, 1769–70 (1996).

[59] *See generally* WILLIAM N. ESKRIDGE, JR., ABBE R. GLUCK & VICTORIA F. NOURSE, STATUTES, REGULATIONS, AND INTERPRETATION: LEGISLATION AND ADMINISTRATION IN THE REPUBLIC OF STATUTES 449–54 (2014); Antonin Scalia, *Common-Law Courts in a Civil-Law System: The Role of United States Federal Courts in Interpreting the Constitution and Laws*, *in* A MATTER OF INTERPRETATION 3, 88 (Amy Gutmann ed., 1997).

All this said, the sufficient motivation for this Article's study of ordinary meaning is that ordinary meaning analysis *is in fact* significant across many legal domains.[60] Across diverse areas of legal interpretation — contracts, wills, trusts, deeds, patents, statutes, regulations, treaties, and constitutions[61] — the ordinary meaning of the legal text is a relevant consideration in determining how it should be applied.

One of the key features of the pure ordinary meaning approach is that it differs in some ways from approaches seeking to determine the drafter's intent. As Justice Holmes notes in the passage above, ordinary meaning analysis considers the understanding of most ordinary people, not the specific intentions of the document's author.[62] Nor should we be concerned with some more normative notion of meaning — what the words in the text should ideally mean. Similarly, the "original public meaning" literature makes clear that public meaning is what the law communicates to its actual audience, which is not necessarily what the drafters aimed to accomplish in drafting the laws.[63]

Of course, even in interpretation seeking to uncover intent, ordinary meaning is often taken to be evidence of intent — and even the only permissible evidence of intent insofar as the ordinary meaning is plain. For example, in the interpretation of wills, intention is the "controlling consideration in determining the meaning,"[64] and most courts use plain meaning rules to establish this meaning, excluding extrinsic evidence when the will's meaning is plain.[65] Although testator intent is the primary interpretive criterion, evidence that the testator intended another meaning cannot disturb the intent indicated by the text's plain meaning.[66]

Sometimes interpreters seek to determine ordinary meaning at a particular time. For example, "original" in original public meaning refers to the time of the text's passage or ratification. The original public meaning of a text is the ordinary or public meaning at the time the text became law. Public Meaning Originalism "seeks to determine 'the meaning the words and phrases of the Constitution would have had, in

---

60 *See* Holmes, *supra* note 52, at 417; *see also supra* pp. 1–2.

61 *See supra* notes 2–10 and accompanying text.

62 Holmes, *supra* note 52, at 419–20.

63 *See generally* Solum, *supra* note 10; *see also* James C. Phillips, Daniel M. Ortner & Thomas R. Lee, *Corpus Linguistics & Original Public Meaning: A New Tool to Make Originalism More Empirical*, 126 YALE L.J.F. 21, 22 (2016).

64 RESTATEMENT (THIRD) OF PROP.: WILLS AND OTHER DONATIVE TRANSFERS § 10.1 (AM. L. INST. 2003).

65 JESSE DUKEMINIER & ROBERT H. SITKOFF, WILLS, TRUSTS, AND ESTATES 328 (9th ed. 2013).

66 *Id.*

context, to ordinary readers, speakers, and writers of the English language, reading a document of this type, at the time adopted.'"[67] In most cases, the text's communicative content is simply its ordinary meaning.[68] Original public meaning is conceptualized as "the likely original understanding of the text at the time of its adoption by competent speakers of the English language"[69] or "what readers of the historically-situated text would have understood the constitutional language to express."[70]

Despite continuing debate about the precise contours of ordinary meaning, there is actually remarkable consistency on this point: ordinary meaning is generally informed by considerations of *how readers of the text would actually understand it*. That is, whether the determination of ordinary meaning is a question of law or fact,[71] it is informed by ordinary understanding of language — either through linguistic intuitions of the judge or jury or sources of evidence about ordinary understanding, like dictionaries or patterns of word usage across corpora.[72]

B. Theories of Interpretation

Take the best-known hypothetical in legal interpretation: "[N]o vehicles in the park."[73] This example illustrates the complexities of legal interpretation, but also the claims of various legal-interpretive theories, such as textualism, intentionalism, and purposivism.[74]

To provide a brief example, we can elaborate the hypothetical with a few more details.[75] Imagine two legal texts that concern Mr. Hart. The first is a modern (2020) insurance contract, which provides Mr. Hart with liability insurance for "covered vehicles," which are defined in the contract as "vehicles owned by you [Mr. Hart], for which no other insurance policy provides coverage." As a second example, imagine that a 1958 "East Dakota" statute requires that "every owner of a vehicle,

---

[67] Phillips et al., *supra* note 63, at 21–22 (quoting Vasan Kesavan & Michael Stokes Paulsen, *The Interpretive Force of the Constitution's Secret Drafting History*, 91 GEO. L.J. 1113, 1118 (2003)).

[68] Lee & Mouritsen, *supra* note 14, at 792. Justice Lee and Professor Mouritsen argue for this equivalence, except in the case of specialized legal language (for example, "bill of attainder" or "parol evidence"). Part IV's experiments and this Article's arguments set aside study of specialized legal language. As such, following Lee and Mouritsen, I treat ordinary meaning to be equivalent to a text's communicative content or public meaning.

[69] KURT T. LASH, THE FOURTEENTH AMENDMENT AND THE PRIVILEGES AND IMMUNITIES OF AMERICAN CITIZENSHIP 277 (2014).

[70] Christopher R. Green, *The Original Sense of the (Equal) Protection Clause: Pre-enactment History*, 19 GEO. MASON U. C.R. L.J. 1, 12 (2008).

[71] Omri Ben-Shahar & Lior Jacob Strahilevitz, *Interpreting Contracts via Surveys and Experiments*, 92 N.Y.U. L. REV. 1753, 1765 (2017).

[72] *See id.*; Macleod, *supra* note 54, at 985.

[73] *See* H.L.A. Hart, *Positivism and the Separation of Law and Morals*, 71 HARV. L. REV. 593, 607 (1958); *see also* Lee & Mouritsen, *supra* note 14, at 800.

[74] *See generally* ESKRIDGE, *supra* note 17.

[75] For a much more detailed analysis, see generally *id.*

before any such vehicle is operated in this state, shall apply for and obtain registration in this state."

We can imagine different legal disputes arising today concerning various of Mr. Hart's possessions: his car, bicycle, and airplane. Mr. Hart leads an adventurous lifestyle, and he has operated all three of these without registration in East Dakota. He also has a streak of bad luck; his operation of the car, bicycle, and airplane each results in a separate accident. No other insurance contract provides coverage for his car, bicycle, or airplane. Does Mr. Hart's insurance contract cover any of these, and for which entities must he have obtained registration?

On various theories of contractual and statutory interpretation, answering both of these questions would involve analyzing the ordinary meaning of relevant provisions; in these cases, that would certainly begin with the ordinary meaning of the term "vehicles."[76] To determine whether Mr. Hart's bicycle or airplane is covered, a formalist approach to contract law might consider just the ordinary meaning of the contract, excluding extrinsic evidence absent a finding of ambiguity. For example, a court adopting a Plain Meaning Rule would not consider evidence of the parties' prior negotiations, insofar as the contract's text is "plain" or "unambiguous."[77]

Textualist theories of statutory and constitutional interpretation would similarly look to the ordinary meaning of "vehicle" in the East Dakota statute. Thirty years ago, Justice Scalia introduced the interpretive theory of "new textualism," noting that "[e]very issue of law resolved by a federal judge involves interpretation of text — the text of a regulation, or of a statute, or of the Constitution."[78] The rise of "new textualism" — broadly speaking, the theory that plain and clear text is decisive of legal effect — came alongside the rise of "new originalism" — broadly speaking, the idea that original public meaning (and not drafters' intentions) constrains interpretation. These views have fused into a modern thesis of Public Meaning Originalism, a thesis inspiring originalist and textualist views in both constitutional and statutory interpretation,[79] but which might also be applied in private law contexts.[80] Unsurprisingly, textualist and originalist theories place great significance on the (original) public meaning of the text. On the most popular

---

[76] *Compare, e.g.*, McBoyle v. United States, 283 U.S. 25, 26–27 (1931) (stating that an airplane is not a vehicle within the meaning of the criminal statute), *with* McReynolds v. Mun. Ct. of Ottumwa, 207 N.W.2d 792, 795 (Iowa 1973) (stating that an aircraft is a vehicle).

[77] Eric A. Posner, Essay, *The Parol Evidence Rule, the Plain Meaning Rule, and the Principles of Contractual Interpretation*, 146 U. PA. L. REV. 533, 534 (1998).

[78] Scalia, *supra* note 59, at 13–14.

[79] *See generally, e.g.*, Solum, *supra* note 10; Lee & Mouritsen, *supra* note 14.

[80] *See generally, e.g.*, Chiang & Solum, *supra* note 22 (on patents); Mouritsen, *supra* note 16 (on contracts).

version of these theories, the original public meaning of legal text constrains the text's effect.[81]

So in this example, a textualist-originalist or "new originalist" view would not look to the 2020 ordinary meaning of "vehicle," but rather to the 1958 ordinary meaning of "vehicle." On that view, the ordinary or public meaning of "vehicle" in 1958 fixes the meaning of the statute and constrains how it applies today. Looking to historical meaning will often make a difference, as meaning can change over time.[82]

Ordinary meaning is a significant interpretive criterion to a range of other legal-interpretive theories — across domains from constitutional and treaty interpretation to the interpretation of trusts and wills. Pluralist theories might take ordinary meaning to be one of several relevant considerations in interpretation.[83] Even if ordinary meaning does not necessarily constrain legal effect — for example, it might be overridden by considerations about intentions of the contracting parties, the statute's purpose, or efficiency and consequences — it still plays a role as an important consideration in legal decisionmaking on many plausible theories.

The politicization of originalism and textualism can obscure widespread agreement about this point. On *many* theories of legal interpretation, how a text is understood by ordinary people is one relevant consideration in determining the text's legal effect. So while the stakes of the present project are highest for textualist theories of interpretation, as well as for views advocating strong plain meaning and exclusive parol evidence rules, the project is also relevant to any theory of legal interpretation that places any significance on ordinary meaning.[84]

It is important to recall that in some circumstances — such as in the interpretation of contracts and wills — interpretive theories justify the ordinary meaning approach as a means of determining intent. That is, the "intention of the parties should control, and the best evidence of intent is the contract itself."[85] On these views ordinary meaning analysis is still concerned with the intention of the parties "as expressed in the

---

[81] Solum, *supra* note 10, at 3.

[82] *Id.*

[83] *See, e.g.*, Stephen M. Griffin, *Pluralism in Constitutional Interpretation*, 72 TEX. L. REV. 1753, 1753 (1994).

[84] Of course, it is conceptually possible that a theory of interpretation does not place *any* interpretive significance on ordinary meaning as either a criterion in itself or evidence of another criterion, in any context. For example, a strong purposivist theory might hold that what *always* determines a text's legal effect is simply the motivating purpose of the text and that the ordinary meaning of that text is never relevant in determining its purpose. Or a strong intentionalist theory might hold that what always determines the legal effect is simply the mutual intention of the drafters or contracting parties and that ordinary meaning is never relevant in determining such intent.

[85] Gary Friedrich Enters., LLC v. Marvel Characters, Inc., 716 F.3d 302, 313 (2d Cir. 2013) (quoting Cont'l Ins. Co. v. Atl. Cas. Ins. Co., 603 F.3d 169, 180 (2d Cir. 2010) (alterations omitted)).

clear language of the contract."[86] To do this, these views often recommend considering how the meaning would be understood "objectively,"[87] by "the average man on the street."[88] In the majority of jurisdictions, if the plain meaning of the contractual language can be established, the court will not look further to other extrinsic evidence of intent.[89] To establish this meaning, courts often cite dictionaries and patterns of word usage.[90] Here again, ordinary meaning is importantly informed by the insights of empirical tools, aiming to reflect a fact about how language is understood.[91]

The next section turns to the question of *how* to determine ordinary meaning, continuing with the illustrative example of "vehicles." It is worth recalling that this example is not of purely philosophical or academic interest. It arose in the well-known case *McBoyle v. United States*,[92] in which Justice Holmes's interpretation of a criminal statute concluded that an airplane is not a vehicle in the ordinary sense of "everyday speech."[93] But it also arises in criminal and insurance law; for example, courts have considered whether a bicycle is a "vehicle" within the meaning of a plea agreement,[94] and whether entities like watercrafts, airplanes, and motorcycles are "vehicles" within the meaning of insurance contracts.[95] For continuity with the theoretical literature, this Article uses as its leading example the question of whether entities like airplanes and bicycles are "vehicles." But as Part IV's experimental

---

86 Parks Real Est. Purchasing Grp. v. St. Paul Fire & Marine Ins. Co., 472 F.3d 33, 42 (2d Cir. 2006) (quoting Morgan Stanley Grp. Inc. v. New Eng. Ins. Co., 225 F.3d 270, 275 (2d Cir. 2000)).

87 *See* Sayers v. Rochester Tel. Corp. Supplemental Mgmt. Pension Plan, 7 F.3d 1091, 1095 (2d Cir. 1993).

88 Lachs v. Fid. & Cas. Co. of N.Y., 118 N.E.2d 555, 558 (N.Y. 1954).

89 *See* Peter Linzer, *The Comfort of Certainty: Plain Meaning and the Parol Evidence Rule*, 71 FORDHAM L. REV. 799, 800 (citing 5 MARGARET N. KNIFFIN, CORBIN ON CONTRACTS § 24.7, at 34 (Joseph M. Perillo ed., 1998)). This is despite rejection of the plain meaning rule by both the Uniform Commercial Code and the Restatement (Second) of Contracts section 212. *See* Linzer, *supra*, at 827–28.

90 *See* Mouritsen, *supra* note 16, at 1340–41.

91 It is worth noting some tension here between the *legal* interpretive question of the text's plain meaning and the seemingly *factual* inquiry conducted with dictionary definitions and corpus linguistics analyses.

92 283 U.S. 25 (1931).

93 *Id.* at 26. *But see* McReynolds v. Mun. Ct. of Ottuma, 207 N.W.2d 792, 796 (Iowa 1973).

94 *See* Anthony v. State, 329 P.3d 1027, 1029 (Alaska Ct. App. 2014) (applying contract law interpretive rules to interpret a plea agreement concerning whether "vehicle" includes a bicycle with an "after-market motor" attached).

95 *See, e.g.*, GEICO Marine Ins. Co. v. Great N. Ins. Co., No. 16-CV-1788, 2017 WL 4286394, at *5–6 (S.D.N.Y. Sept. 11, 2017) (watercraft); Certain Brit. Underwriters at Lloyds of London v. Jet Charter Serv., Inc., 789 F.2d 1534, 1535 (11th Cir. 1986) (airplanes); Marasco v. Hopewell, No. 03AP-1081, 2004 WL 2895973, at *6–8 (Ohio Ct. App. Dec. 14, 2004) (motorcycles).

work demonstrates, the scope of the Article's argument extends to the interpretation of many other terms and phrases.[96]

II. SOURCES OF INTERPRETIVE EVIDENCE

How do legal interpreters find the "ordinary meaning" of a text? Different sources of evidence bear on this question. This Part outlines three of the most popular sources of interpretive evidence: individual intuition, dictionary definitions, and patterns of word usage, as revealed by legal corpus linguistics methods.

A. Individual Intuition

A common source of interpretive evidence is individual intuition.[97] What does the ordinary meaning of the contract or will *seem* to be; or what does it *seem* was the original public meaning of the statute or constitution?

Ordinary people's *collective* understanding of legal texts is closely connected to — and on some views, constitutive of — ordinary meaning.[98] However, it is crucial to distinguish between collective intuition and individual intuition. Typically when judges and scholars employ intuition in interpretation, they do not refer to survey evidence or panels of ordinary speakers. Rather, they rely on their own individual intuitions.

In easy cases, linguistic intuition is often a helpful source in identifying ordinary meaning. Consider again the well-known example of a local ordinance prohibiting "vehicles" from entering a park.[99] Most people today would understand that the ordinance would not prohibit someone from bringing their baby in a baby shoulder-carrier into the park *because* each of us understands that a baby shoulder-carrier is not a vehicle. Even if there is some disagreement about some entities (for example, is a skateboard a vehicle?), linguistic intuition provides straightforward guidance in many other cases (for example, baby shoulder-carriers may be brought into the park; cars must be kept out).

In harder cases, however, linguistic intuition may not helpfully identify ordinary meaning — particularly when there is substantial disagreement. We should not expect one person's linguistic intuition to necessarily track ordinary meaning. In fact, research suggests that people often are subject to a false consensus bias, thinking (incorrectly) that they are good measures of the population's consensus.[100] This concern is amplified by the fact that it is not always clear how an individual

---

[96] Specifically, the account here concerns any term or phrase that admits of competing "prototypical" and "extensive" interpretations. As Part IV suggests, many legal interpretive debates involve such terms, including "vehicle," but also "labor," "weapon," "tangible object," and even phrases like "carrying a firearm" and "using a firearm." *See infra* Part IV, pp. 753–77.

[97] *See* Mouritsen, *supra* note 16, at 1346–48.

[98] One reason to reject the constitutive claim is that individuals sometimes make errors.

[99] This is discussed in much greater detail as the leading example in Parts III–V.

[100] *See* Solan & Gales, *supra* note 44, at 1333; Lawrence Solan, Terri Rosenblatt & Daniel Osherson, *False Consensus Bias in Contract Interpretation*, 108 COLUM. L. REV. 1268, 1268–69 (2008).

intuiter could know whether she faces an easy or hard case. Is one's confidence in individual intuition a reflection of facing a truly easy case, or false consensus bias?

This concern is perhaps most salient in historical interpretation — where our individual intuitions might mistakenly reflect or import the *current* ordinary meaning rather than the *original* one. As thoughtful originalists have cautioned, in the search for original public meaning, "linguistic intuitions formed by immersion in modern linguistic practices can be misleading."[101] Consider Professor Lawrence Solum's astute observation about the Seventh Amendment's "Twenty Dollars Clause." Most modern readers would assume that this clause refers to the modern Federal Reserve note.[102] But, writes Solum, "the word 'dollar' almost certainly referred to the Spanish silver dollar . . . . The 'greenback,' a precursor to the modern note, was not created until much later and was the subject of much controversy."[103]

Individual intuition is recognized — by originalists and nonoriginalists alike — as a fallible source of evidence in modern interpretation and a highly dubious source of evidence in historical interpretation. Intuition's errors can also be hard to identify in practice. Linguistic intuition often *feels* very compelling: surely "dollar" in the Constitution means dollar. Sometimes, historical research shakes linguistic intuitions of this misplaced confidence. But in legal interpretation, it is a dangerous strategy to rely on unreliable linguistic intuitions *until* and *unless* they are proven erroneous.

There is also an important question of whether individual intuition-use is even a true *method* of interpretation. While we can generate principles to guide the use of dictionaries and corpus linguistics, it is less clear how judges should practice "individual intuition." Given the errors that individual intuition can produce in interpretation, such a method requires some guiding principles.

B. Dictionary Definitions

Dictionaries have prima facie plausibility as useful sources in interpretation. After all, if the aim is to discover the (original) ordinary meaning of a term, why not look at how the relevant community defines that term? Recall our example of "vehicles." One might seek evidence about the meaning of "vehicle" in 1958 by considering a dictionary definition from that time. As Justice Thomas Lee and Professor Stephen

---

101 Lawrence B. Solum, *Originalist Methodology*, 84 U. CHI. L. REV. 269, 281 (2017).

102 *Id.* at 281–82. This assumption also arises in some legal scholarship. *See, e.g.*, Note, *The Twenty Dollars Clause*, 118 HARV. L. REV. 1665 (2005).

103 Solum, *supra* note 101, at 282.

Mouritsen note, *Webster's Third New International Dictionary* (1961) defines vehicle as a "means of carrying or transporting something."[104]

Part IV presents experimental work that provides evidence about the reliability of dictionaries in interpretation. But there are other aspects of dictionary use that call for analysis and critique.[105] For one, dictionary definitions may be *normative*. That is, while one might look to a dictionary for evidence about how some term was — in fact — understood, a dictionary may instead report the normative view of its author(s) concerning how some term *should have been* understood. If so, in at least some cases, dictionary definitions are not tracking ordinary meaning. Rather, they might be tracking the dictionary drafter's conception of desirable meaning. Often, legal interpretive disputes turn on questions about subtle shades of meaning, so such a difference could be consequential.

Another concern involves the possibility of "dictionary shopping."[106] A single term may be defined several ways within a single dictionary and across multiple dictionaries. Without any principled method to choose among definitions and dictionaries, this opens the possibility that interpreters might sometimes choose definitions that suit their political or personal preferences with respect to the case at hand.

In the search for historical meaning, another important limitation is that historical dictionaries were published less frequently than modern ones are. For example, two leading English language dictionaries were published seventy-three years apart, in 1755 and 1828.[107] Although language change is usually gradual, there are obvious questions about the limits of these dictionaries in interpreting the Constitution and early amendments. How reliable is a 1755 dictionary in reflecting the meaning of a provision from 1789? Relatedly, it is not clear that historical dictionaries provide accurate or complete portraits of language in use at the time of their publication.[108] Here, too, this raises obvious questions about the reliability of these sources. We might cross-check historical dictionaries to illuminate idiosyncrasies, but the limited number of historical dictionaries severely limits the usefulness of this effort.

Despite all of these concerns, dictionaries are an increasingly popular source of interpretive evidence, at least among the Justices of the

---

104 Lee & Mouritsen, *supra* note 14, at 801.

105 *See* James J. Brudney & Lawrence Baum, *Oasis or Mirage: The Supreme Court's Thirst for Dictionaries in the Rehnquist and Roberts Eras*, 55 WM. & MARY L. REV. 483, 492–93 (2013) (summarizing the central issues related to judicial reliance on dictionaries); Solan & Gales, *supra* note 44, at 1334 (same).

106 *See, e.g.*, Ellen P. Aprill, *The Law of the Word: Dictionary Shopping in the Supreme Court*, 30 ARIZ. ST. L.J. 275, 281 (1998).

107 Solum, *supra* note 33, at 246.

108 *Id.*

Supreme Court.[109] And dictionaries' use is not without some initial plausibility: it is reasonable to hypothesize that the ordinary meaning of a term is reflected well by its dictionary definition. Whether this is true is an open empirical question.

C. Linguistic Usage Data via "Legal Corpus Linguistics"

A final source of interpretive evidence comes from "legal corpus linguistics." Corpora are sets of language data, containing text from books, newspaper articles, online publications, and other sources.[110] In recent years, legal corpus linguistics has evolved from smaller searching to a "big data" approach.[111]

The most prominent defense of this new legal corpus linguistics approach is Lee and Mouritsen's *Judging Ordinary Meaning*.[112] That paper advocates a promising account of corpus linguistics use in determining the ordinary meaning of legal texts, and defends corpus linguistics as an objective, scientific, data-driven approach to legal interpretation.[113]

The core of Lee and Mouritsen's analysis involves two types of corpus searches: "collocation" and "keywords in context." A "collocation" search in a corpus shows the words that are most likely to appear in the same context as the search term.[114] A "keywords in context" search presents the user of corpus linguistics with examples of the term in context.[115]

Take the "no vehicles in the park" example. One might seek evidence about the ordinary meaning of "vehicle" in 1958 by considering data from the corpus at the time. What are the common collocates of "vehicle"; with what other words is "vehicle" typically used in the corpus?

Lee and Mouritsen provide a corpus analysis of this exact question. They maintain that collocation provides "a snapshot of the semantic environment in which *vehicle* appears and the kinds of vehicles that tend to appear in that environment."[116] For example, in a modern search,

---

[109] Calhoun, *supra* note 39, at 497.

[110] *See generally* DOUGLAS BIBER, SUSAN CONRAD & RANDI REPPEN, CORPUS LINGUISTICS: INVESTIGATING LANGUAGE STRUCTURE AND USE (1998) (setting forth a framework for corpus linguistics and definition of corpora).

[111] Solan & Gales, *supra* note 44, at 1336. The past few years have seen increased scholarly interest in corpus linguistics. *See generally* Carissa Byrne Hessick, *Corpus Linguistics and the Criminal Law*, 2017 BYU L. REV. 1503; Neal Goldfarb, *A Lawyer's Introduction to Meaning in the Framework of Corpus Linguistics*, 2017 BYU L. REV. 1359; Stefan Th. Gries & Brian G. Slocum, *Ordinary Meaning and Corpus Linguistics*, 2017 BYU L. REV. 1417; Recent Case, Wilson v. Safelite Group, Inc., *930 F.3d 429 (6th Cir. 2019)*, 133 HARV. L. REV. 691 (2019).

[112] *See generally* Lee & Mouritsen, *supra* note 14.

[113] *See id.*

[114] *Id.* at 832.

[115] *Id.*

[116] *Id.* at 837.

the top collocates of "vehicle" include "electric," "motor," "gas," "autonomous," and so on. As Lee and Mouritsen infer, "[m]any of the collocates of *vehicle* . . . strongly indicate *automobile* as a likely candidate for the most common use of the term. . . . *Airplane* does not appear . . . . Similarly, *bicycle* does not appear among the collocates of *vehicle* in contemporary usage."[117]

Next they conduct a keywords in context (or concordance line) search. This search returns examples of the use of "vehicle" in context. For example, "the driver . . . apparently lost control of the vehicle because he was traveling too fast for the wet road conditions."[118]

It is important to note that this popularized use of *legal corpus linguistics* — focused on collocation and keywords in context — is very different from other versions of corpus linguistics, including ones that attempt to build more complex statistical or computation models of meaning, or those that use algorithmic processes (for example, using word2vec to analyze word embeddings). Practicality concerns likely motivate this difference. This simpler "law and corpus linguistics" method — in which a human evaluates patterns of common language use across a range of written sources — is one that many legal interpreters (for example, judges) can employ cheaply and swiftly. And they have.[119] A legal corpus linguistics "revolution" imagines judges, without much additional technical training, running searches like collocation and keywords in context to assess the frequency of usage in corpora.

This Article focuses on this current form of legal corpus linguistics, which examines common usage via searches like collocation and keywords in context. The Article's critical conclusions, including the burden-shifting argument, invite responses from advocates of these and other methods. As such, the Article remains open to the possibility that some other approach that might fall under the broad umbrella of "corpus linguistics" may answer some of the problems identified here. Of course, the most successful reply would not simply gesture to some different technical possibility (for example, consider "dispersion," or "use word2vec"). Rather, the successful reply would articulate how the method should be used and when, and also demonstrate *empirically* that the method, in fact, reliably reflects ordinary meaning.

III. ORDINARY MEANING METHODOLOGY'S EMPIRICAL ASSUMPTIONS

When using an empirical method to provide evidence about ordinary meaning, like assessing dictionary definitions or linguistic usage data, there are a number of possible empirical critiques. Perhaps the most

---

117 *Id.* at 837–38.
118 *Id.* at 841.
119 *See* sources cited *supra* note 37.

central critiques concern accuracy and reliability: Does the use of a dictionary lead interpreters to accurate verdicts regarding ordinary meaning (for example, in the case at hand), and does it reliably lead interpreters to correct verdicts (for instance, across the set of cases in which the method is used)? But there are a host of other related critiques: Is there inconsistency among different users of the same method; does the method lead interpreters to verdicts inconsistent with the recommendations of another (accurate or supposedly accurate) method; is the method plagued by arbitrary choices? Section A articulates these critiques. The Article's focus remains on the central accuracy and reliability critiques, as well as inconsistency concerns, but various other critiques remain relevant when assessing dictionary and legal corpus linguistics use.[120]

With these empirical questions in mind, section B turns a critical eye to the use of dictionaries and legal corpus linguistics in interpretation. It poses a thought experiment: What might future interpreters glean from modern dictionaries and legal corpus linguistics analyses in interpreting a modern legal text? This hypothetical sheds light on what dictionaries and legal corpus linguistics might commonly suggest to interpreters about ordinary meaning. Of course, this thought experiment is merely suggestive. In Part IV the Article presents a more rigorous examination of the hypotheses generated from the thought experiment.

A. Empirical Critiques

The table below outlines different empirical critiques and examples of each critique leveled at one source of evidence (for example, individual intuition, dictionary use, or legal corpus linguistics). Most of these are "internal" empirical critiques. That is, they do not criticize interpretive theories and methods from an external perspective, asking whether we *should* be committed to applying a text in line with its ordinary meaning. Instead, they critique these theories on their own terms, asking whether we *can* achieve the task they set: Do the task's *own* empirical assumptions withstand scrutiny?

---

[120] Moreover, the empirical results of Part IV might also be taken to support various of these other critiques. Much of that analysis falls outside the scope of the present Article, which deals primarily with the central accuracy critique and the inconsistency critiques.

Table 1: Empirical Critiques of Originalist Methods

Critique	Example of Claim
Accuracy: The method recommends a false verdict concerning ordinary meaning.	Some method (for example, dictionary use) recommends an interpretive result that is inconsistent with ordinary meaning.
Reliability: The method is not a reliable measure of ordinary meaning.	Some method (for example, corpus linguistics) does not systematically track the criterion/criteria of ordinary meaning.
Inconsistency Among Interpreters: Different people using the method reach different verdicts.	Different judges have different linguistic intuitions about ordinary meaning.
Inconsistency Within Interpreters: The same person using the method reaches different verdicts.	A judge's use of corpus linguistics at one time or in one context recommends an interpretive result that is inconsistent with the result recommended by the same judge's use of corpus linguistics at some other time or in some other sufficiently similar context.
Inconsistency with Other Methods: One method recommends a different verdict from that recommended by others.	An interpreter's individual linguistic intuition conflicts with the recommendation generated by their use of dictionaries or corpus linguistics.
Inconsistency Within a Method: The method provides evidence for divergent verdicts.	Definitions from two dictionaries provide divergent recommendations about ordinary meaning; two different plausible search criteria of corpus linguistics provide divergent recommendations about ordinary meaning.
Arbitrariness in Practice: The method's current actual use is plagued by arbitrary decisions.	There is no principled application of legal corpus linguistics, as it is currently used in ordinary meaning interpretation.
Arbitrariness in Theory: The method's use cannot escape arbitrary decisions.	There is no principled decision among conflicting dictionaries or alternate definitions.
Interpretive Underspecificity: Use of the method does not provide determinate outcomes.	In a corpus linguistics data set, what frequency of use indicates that the use is part of the "ordinary meaning"?

Interpretive Vagueness: Assumptions required to use the method admit of problematic vagueness or implausible cutoffs.	A theory holds that if a particular use of a term is reflected in less than 5% of the corpus, then that use is not part of the ordinary meaning of the term. But why not 3%; why not 10%?
Bias: Use of the method enables political values or bias to influence interpretation.	Individual intuition is subtly or unconsciously influenced by politically motivated reasoning; interpreters "cherry-pick" definitions that support the interpretation consistent with their political beliefs.
Impracticality: The method is too complicated, expensive, or otherwise impractical to use.	Some forms of corpus linguistics require technical training that judges do not have.

These critiques can each be posed for different methods: individual intuition, dictionary use, and corpus linguistics. This Article focuses primarily on the Accuracy and Reliability critiques for dictionary use and corpus linguistics, and to a lesser extent the Inconsistency critiques. However, it is important to note that there are many other critiques to be considered in assessing the use of dictionaries and corpus linguistics in ordinary meaning analysis.

B. Are Dictionaries and Legal Corpus Linguistics Reliable? A Thought Experiment

With the increasing use of both dictionaries and legal corpus linguistics in interpretation, a crucial question looms: Are these methods actually achieving their aims? Often, interpreters simply assume that dictionary use and legal corpus linguistics reflect facts about ordinary meaning.[121] But this is an open empirical question.

The question has remained open, perhaps, because it seems untestable. To know whether these methods are reliable, we need some verification source, a Rosetta Stone of truths about ordinary meaning, or *original* ordinary meaning. To know whether an eighteenth-century dictionary reflects the ordinary meaning of the time, we need true facts about ordinary meaning in the relevant context of the eighteenth century. The fact that we lack access to any direct verification source motivates interpreters to use other methods (like dictionaries and legal corpus linguistics) to provide evidence about ordinary meaning.

---

[121] *See, e.g.*, Brudney & Baum, *supra* note 105, at 486–87, 568; Gries & Slocum, *supra* note 111, at 1441.

But perhaps a Rosetta Stone of ordinary meaning is not the only option. Although we lack precise verification about historical (for example, 1787) ordinary meaning, we are much more confident in modern ordinary meaning. Our familiarity with modern meaning can help assess sources of historical interpretive evidence.

Consider a (toy) thought experiment. Imagine that a modern constitutional amendment stated that "vehicles" must be registered with the federal government. Two hundred years later, in 2220, a legal dispute erupts concerning the 2020 original public meaning of "vehicle." Would consulting 2020 dictionaries and legal corpus linguistics provide precise and reliable evidence about the amendment's original meaning?

Following Lee and Mouritsen's recent defense of originalist and textualist methodology, consider their example of a dictionary definition of a vehicle: "a means of carrying or transporting something," or an "agent of transmission" or "carrier."[122] An interpreter in 2220 who uses this definition might think that roller skates, or zip lines, or even baby-shoulder carriers are vehicles. But people today generally judge that these entities are not vehicles.[123] The amendment's 2020 ordinary meaning is not that roller skates must be registered. But relying on the dictionary would suggest precisely the opposite.

Legal corpus linguistics may fare no better. As Lee and Mouritsen note, the written word "vehicle" almost always refers to a car.[124] And it most often appears near words associated with cars, like "electric" and "motor."[125] This reflects one common use of "vehicle," but it neglects other acceptable uses. We do not often write today about horse-drawn carriages as "vehicles," and they aren't described as having "motors" or "electric" power. But we understand that they are vehicles. Legal corpus linguistics might suggest that airplanes and helicopters are not "vehicles." But it is far from obvious that the ordinary meaning of the modern amendment excludes those entities.

This thought experiment raises questions about the accuracy of dictionaries and legal corpus linguistics. The worry is not that these methods get things wrong in some unusual or esoteric cases. Insofar as legal corpus linguistics suggests that airplanes are not part of the ordinary meaning of "vehicle," this is a mistake about a very common term — and one that has been litigated more than once.[126] This reflection should

---

[122] Lee & Mouritsen, *supra* note 14, at 800–01 (citing WEBSTER'S THIRD NEW INTERNATIONAL DICTIONARY 2538 (1961)) ("[O]ne attested sense of *vehicle* is the notion of a 'carrier' or 'agent of transmission.'"). Lee and Mouritsen hypothesize that this sense of "vehicle" could "sweep broadly." *Id.*

[123] For empirical evidence of this claim, see *infra* Figure 5.

[124] Lee & Mouritsen, *supra* note 14, at 801.

[125] *Id.* at 837.

[126] *See, e.g.*, McBoyle v. United States, 283 U.S. 25 (1931); McReynolds v. Mun. Ct. of Ottumwa, 207 N.W.2d 792 (Iowa 1973).

give interpreters pause when using dictionaries or legal corpus linguistics, especially when relying on just one of these methods as the sole source of interpretive evidence.

The argument requires one more step in the historical context. The central empirical assumption of views like Public Meaning Originalism is that its tools (such as dictionary use and legal corpus linguistics) reflect original meaning. This assumption remains surprisingly underexplored,[127] and the thought experiment suggests that it may not be true. If people's modern judgments are not reflected by a method's modern use, we can argue by a historical inference that the method is unreliable in historical (originalist) interpretation:

1. Empirical Claim: The modern use of a method (for example, use of dictionaries or legal corpus linguistics) does not accurately reflect people's ordinary judgments.
2. Reliability Premise: A method that does not accurately reflect people's judgments is not a reliable method of determining ordinary meaning.
3. Intermediate Conclusion: There is evidence that the method is unreliable in modern interpretation.
4. Historical Inference: In the absence of historically distinguishing factors, evidence of a method's unreliability in modern interpretation also serves as evidence about that method's unreliability in historical interpretation.
5. Conclusion: There is evidence that the method is unreliable in historical interpretation.

An important piece of this argument to unpack is the historical inference. This premise holds that in the absence of historically distinguishing factors, evidence of a method's modern unreliability is also evidence of that method's historical unreliability. A "historically distinguishing factor" would be a compelling reason to think that use of a method is more reliable in historical interpretation. In the case of dictionaries and legal corpus linguistics, most of the factors pull in the opposite direction. Modern dictionaries are larger and more frequently revised. Modern corpora are vastly larger and far more easily searchable than historical corpora. Finally, *modern* use of a tool is presumably at least as accurate in reflecting modern ordinary meaning as is *modern* use of a historical tool in reflecting historical public meaning.

Of course, in both the modern and historical versions of the argument, another crucial premise one might question is the empirical claim.

---

[127] There has been some prior empirical research on originalism. However, these studies address different questions from those considered here. For example, in an important study, Professor Frank Cross suggests that originalism does not, in fact, effectively restrain willful judging. FRANK B. CROSS, THE FAILED PROMISE OF ORIGINALISM 189 (2013); *see also* sources cited *supra* note 45.

Does this thought experiment really show that these methods do not track modern ordinary meaning? The next Part addresses the empirical claim head-on by presenting an experimental test of what dictionary definitions and linguistic usage data suggest to legal interpreters.

IV. AN EXPERIMENTAL TEST OF DICTIONARIES AND LEGAL CORPUS LINGUISTICS

This Part turns to an experimental assessment of dictionaries and legal corpus linguistics in interpretation. These experiments focus first on the accuracy and reliability of these methods in reflecting ordinary meaning. Subsequent experiments replicate the findings in expert populations (lawyers and U.S. judges) and beyond the example term ("vehicle"). There are four main experiments presented in the main text, and further details can be found in the Appendices.

After sections A through D present the experiments, section E offers a crucial summary and interpretation of the main experimental findings. First, there was a surprising similarity between expert and nonexpert interpreters. Judges, law students, and ordinary people were strikingly similar in their ordinary conceptual judgments, use of dictionaries, and use of legal corpus linguistics data. Second, across the studies, users of legal corpus linguistics tended to identify prototypical examples (for example, a car is a vehicle) better than nonprototypical examples (for example, a moped or airplane is a vehicle). Conversely, users of dictionaries sometimes made very extensive judgments (for example, a pair of roller skates is a vehicle).

Finally, the verdicts of both legal corpus linguistics and dictionary users diverged from the verdicts of ordinary people about simple questions like whether an airplane is a vehicle. Insofar as "ordinary meaning" is an empirical fact, premised on how ordinary people actually understand language, those results tell us something very important about ordinary meaning today. Translating those ordinary judgments into modern ordinary meaning, dictionaries and legal corpus linguistics had a 20–35% divergence rate on average and an 80–100% divergence rate for the hardest examples.

A. Experiment 1

The first experiment tested the verdicts delivered by dictionary and legal corpus linguistics use, as compared to ordinary judgments. To minimize researcher degrees of freedom, this experiment used the first test case mentioned by Lee and Mouritsen's recent article endorsing the use of legal corpus linguistics in legal interpretation.[128] That case is the

[128] Lee & Mouritsen, *supra* note 14, at 800–01.

well-known "no vehicles in the park" example.[129] The experiment also used the exact corpus method used in that article[130] and the dictionary definition of "vehicle" that the article describes.[131] That is, it used the "News on the Web" (NOW) Corpus and collocation and keywords in context searches to generate the corpus data.[132]

The experiment divided participants into three groups: legal corpus linguistics, dictionary, and "ordinary concept" participants. The corpus and dictionary participants received legal corpus linguistics or dictionary data about the term "vehicle," while concept participants received no information so that they would rely on their ordinary understanding. Each participant answered whether each of a series of ten entities was a vehicle.

Method

Participants. Two hundred six participants were recruited from Amazon's Mechanical Turk (52% female, 48% male, 0% nonbinary, mean age = 36). Mechanical Turk (MTurk) is an online platform that enables researchers to collect large samples from a population that is more representative than many other typical research samples.[133]

Materials and Procedure. Participants were randomly divided into one of three conditions: Dictionary, Legal Corpus Linguistics, or Concept. In each condition, participants received some information about a term. Afterwards, participants evaluated whether ten items (for example, car) were members of the category denoted by the term (for example, vehicle).

In the Concept condition, participants were simply asked to consider the noun "vehicle." Then they were asked to categorize ten entities. For

---

129 *Id.*

130 *See id.* at 836–45.

131 *See id.* at 840.

132 Note that Experiment 4 used the other corpus recommended by Lee and Mouritsen: the Corpus of Historical American English (COHA). *See id.* at 835.

133 *See* Adam J. Berinsky et al., *Evaluating Online Labor Markets for Experimental Research: Amazon.com's Mechanical Turk*, 20 POL. ANALYSIS 351, 366 (2012); Gabriele Paolacci et al., *Running Experiments on Amazon Mechanical Turk*, 5 JUDGMENT & DECISION MAKING 411, 412–13 (2010); *see also* Michael Buhrmester, Tracy Kwang & Samuel D. Gosling, *Amazon's Mechanical Turk: A New Source of Inexpensive, Yet High-Quality Data?*, 6 PERSPS. ON PSYCH. SCI. 3, 5 (2011). However, there are notable critiques of MTurk. *See generally* Richard N. Landers & Tara S. Behrend, *An Inconvenient Truth: Arbitrary Distinctions Between Organizational, Mechanical Turk, and Other Convenience Samples*, 8 INDUS. & ORGANIZATIONAL PSYCH. 142, 152–53 (2015) (acknowledging some concerns regarding MTurk); Gabriele Paolacci & Jesse Chandler, *Inside the Turk: Understanding Mechanical Turk as a Participant Pool*, 23 CURRENT DIRECTIONS PSYCH. SCI. 184, 187 (2014). In these experiments, this Article takes the participants from MTurk to be competent users of the English language, at least with respect to ordinary terms like "vehicle." This assumption's plausibility is strengthened when noting the striking similarity in judgments among MTurkers, law students, and U.S. judges. *See infra* section IV.E, pp. 766–77.

example, they were asked: "Is an automobile a vehicle?" [Yes / No]; "Is a car a vehicle?" [Yes / No]; and so on.

In the Dictionary condition, participants were given a dictionary definition of a vehicle[134]:

1) a means of carrying or transporting something
2) an agent of transmission: carrier

However, participants were not told to which term that definition applied. Instead, they were told that the definition applied to a fake term, an "ailac" ("Consider this dictionary definition of 'ailac' (noun):"). This fake term guaranteed that any associations with the term "vehicle" would not interfere with participants' use of the dictionary.[135] To see the necessity of this design, imagine that dictionary participants evaluated "vehicles," not "ailacs." There would be no way to assess whether any success in dictionary use was attributable to use of the *definition* or people's conceptual competence concerning vehicles. This methodology ensures that each condition reflects only the use of one method of analysis — ordinary conceptual competence, dictionary definition, or corpus data.[136]

Legal Corpus Linguistics participants first saw this information:

> Consider the noun, "ailac." To help understand this term, consider some information about the use of "ailac."
>
> First, consider the top common words used in connection with "ailac." These words might appear before or after ailac, or sometimes close to ailac, for example "electric ailac"; "ailac charging"; "drove the ailac"; and so forth.
>
> Top common words: *electric, motor, plug-in, unmanned, armored, connected, cars, aerial, charging, pure, launch, owners, hybrid, traffic, fuel, driving, gas, autonomous, struck, operating, road, safety, accidents, battery, ownership, emergency, batteries, emissions, seat, advanced, driver, primary, demand, commandeered, fuel-efficient,*

---

134 Note that this dictionary definition mirrors the one suggested by Lee and Mouritsen. *See* Lee & Mouritsen, *supra* note 14, at 800 ("One attested sense of *vehicle* is the notion of a 'carrier' or 'agent of transmission.'" (quoting WEBSTER'S THIRD NEW INTERNATIONAL DICTIONARY 2538 (1961))).

135 This is a common method in linguistics. *See, e.g.*, Jean Berko, *The Child's Learning of English Morphology*, 14 WORD 150, 153–58 (1958) (describing the "wug test"). Thanks to Larry Solan and Tammy Gales for this suggestion.

136 All participants received the following introduction to the experiment:
> In the following screen you will see some information about a term. The term might be a real term that you know (for example, a "painter") or one that is made up (for example, a "krob"). If the term is one that is made up, the "information" about the term will also be fictional. After you see the information, we will ask some questions about the term.

automakers, demonstrators, excluding, lunar, passenger, fleet, gasoline, luxury, drove, parking, retirement, infrastructure.

Next, Legal Corpus Linguistics participants saw further examples of the term in context.[137] This corpus data is *precisely* what recent advocates of legal corpus linguistics recommend.[138] Afterwards, participants in the Dictionary and Corpus conditions categorized ten entities. They were asked: "Is a car an ailac?" [Yes / No], and so on.

Results

As predicted, there were significant differences among Dictionary, Legal Corpus Linguistics, and Concept conditions. Figure 1 indicates the proportion responding "yes" for each entity in each condition.

---

[137] "Next, consider some further examples of 'ailac' in context:

1) . . . the driver, Bhaskar Jha, apparently lost control of the **ailac** because he was traveling too fast for the wet road conditions . . .

2) . . . of the troopers. Parrott says the suspects in the **ailac** began showing aggression and shots rang out. Corporal Shane . . .

3) . . . injury and leaving a child under 12 unsupervised in a motor **ailac** but released on a written promise to appear.) Risk . . .

4) . . . Hybrid electric **ailac**s use regenerative braking (when the **ailac** captures energy that would be otherwise lost from braking) and . . .

5) . . . pushed onto the property because of the speed of which these **ailacs** collide,' said Dr. Tom Lawrence, of Clinical Nutrition . . .

6) . . ., 2009. That day the two officers saw an **ailac** connected to a domestic violence case in which shots had been . . .

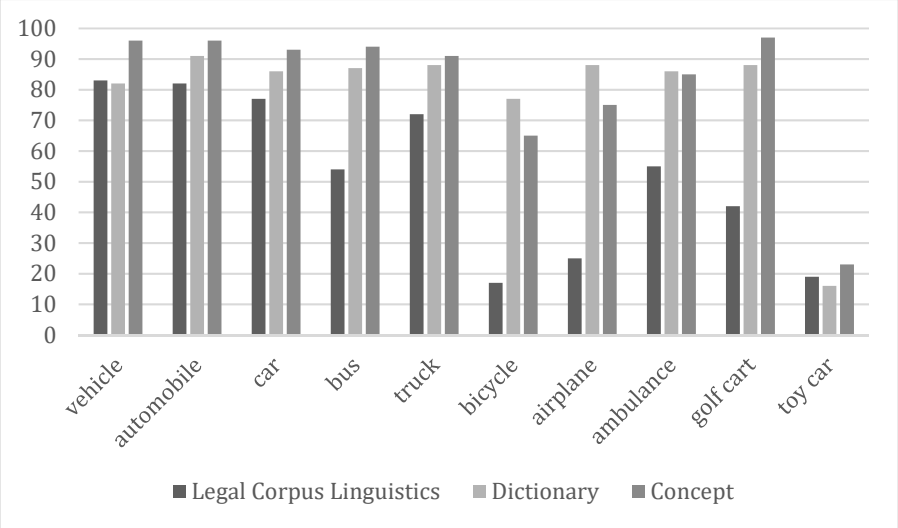
7) . . . say automakers would be better. Wakefield says autonomous **ailacs** could erode the image of certain brands more than others. Brands . . .

8) . . . biogas, and Daimler, which supplies a number of experimental **ailacs** designed to run on natural gas. The German Federal Ministry of . . .

9) . . . is that they aren't kept on file with the Motor **Ailacs** Division or any other entity. By contrast, beneficiary . . ."

[138] Lee & Mouritsen, *supra* note 14, at 840–42.

Figure 1: Percentage Responding "Yes" (the Entity Is a Vehicle) by Dictionary, Legal Corpus Linguistics, and Concept Conditions



To analyze the results, I entered judgment as the dependent measure in a generalized linear mixed-effects model, with participant and entity as random effects, and Method (Corpus, Concept) as fixed effects. There is an effect of Method (Corpus: OR = .10, 95% CI [.05, .20], z = -6.76, p < .0001; Concept: 1.14, 95% CI [.59, 2.18], z = .12, p = .70).

Notably, across the methods, the verdicts differ. Moreover, there is a systematic pattern: Legal Corpus Linguistics participants failed to include a number of entities that are vehicles in the ordinary sense of the term. For example, consider entities like bicycles, airplanes, and golf carts. Although there is some disagreement, most people classify these entities as vehicles. Yet users of legal corpus linguistics largely judged that they are not vehicles.

Discussion

This first experiment represents a small test of the reliability of legal corpus linguistics and dictionaries in reflecting what ordinary people understand language to mean. Broadly speaking, dictionary use was fairly consistent with people's ordinary judgments: cars, buses, and trucks are vehicles, but a toy car definitely is not.

However, legal corpus linguistics did not perform nearly as well. A bus is seemingly within our modern conception of a vehicle, but *only half* of the users of legal corpus linguistics made that categorization. The divergence was not limited to that example. For five of the ten entities, Legal Corpus Linguistics was underinclusive.

One might worry that some participants in the Dictionary or Legal Corpus Linguistics condition were able to guess that "ailac" was substituted for "vehicle." If they had, their inference would suggest that the Concept vs. Dictionary or Concept vs. Legal Corpus Linguistics results may underestimate the true degree of difference. Given the magnitude of the differences among the three conditions, it seems unlikely that all Dictionary or Corpus participants inferred the substituted term; but even if some did, then the results displayed in Figure 1 indicate a *minimum* divergence between ordinary judgment (in the Concept condition) and what a dictionary or legal corpus linguistics search would suggest to interpreters.

A second possible worry is that the "ailac" nonce term suggested to Legal Corpus Linguistics or Dictionary users that there was only one right answer: "ailac" must mean "car" or "bus," but it cannot refer to multiple entities. Although this is an interesting suggestion, it is not consistent with the empirical findings. For both the Legal Corpus Linguistics and Dictionary conditions, over 70% of participants categorized multiple entities as vehicles — for example, vehicle, automobile, car, and truck. Moreover, if participants thought that only one entity was meant to fit the description, categorization percentages should be on average 10%. However, in both conditions, the average was much higher.

B. Experiment 2: The Process of Using Dictionaries and Frequency Data

Consider the patterns of judgment revealed by the previous experiment. Certain entities elicit dramatic differences between Legal Corpus Linguistics and Dictionary participants. For example, the majority of Dictionary participants judged bicycles, airplanes, and golf carts to be vehicles. Yet Corpus participants did not judge these entities to be vehicles.

What explains these differences? Why do legal corpus linguistics verdicts differ from dictionary ones, and why do both sometimes differ from the verdicts supplied by ordinary concept use? One plausible hypothesis draws from research in linguistics and psychology on prototypes.[139] According to prototype theory, people associate concepts with certain features, and more "prototypical" category members are those that have more of those features. For example, both a robin and a penguin are birds, but a robin is a prototypical bird.[140] Experimental studies have shown that people are faster in categorizing prototypical category members than nonprototypical ones. For example, people will

---

[139] *See generally* Eleanor Rosch, *Cognitive Representations of Semantic Categories*, 104 J. EXPERIMENTAL PSYCH. 192 (1975) (describing prototype theory).

[140] *Id.* at 232.

categorize a robin as a bird more quickly than they categorize a penguin as a bird. Moreover, when people are asked to name examples of category members, they cite the more prototypical members more frequently.[141] For example, if you ask people to name a type of pet, they will name "dog" more often than "kangaroo."

I hypothesized that prototype theory might explain some of the differences between dictionaries, which often report broad definitions, and legal corpus linguistics, which reports data indicative of the most frequent and popular uses. As such, it is plausible that legal corpus linguistics data might supply the most useful information about only the more *prototypical* category members. Legal corpus linguistics data provide details about the most frequent uses of a term and the most common words associated with the term. For example, legal corpus linguistics data about vehicles indicate that certain words often appear near "vehicle," such as "motor" and "electric." Perhaps these frequency data are really supplying the most helpful information about *prototypical* category members. This experiment tests that hypothesis.

Method

Participants. One hundred one participants were recruited from Amazon's Mechanical Turk. Eighty-two passed a comprehension-check question (51% female, 48.5% male, 0.5% nonbinary, mean age = 36).

Materials and Procedure. Participants were trained to understand the difference between prototypical and nonprototypical category members.[142] Participants were then instructed to "[c]onsider the noun 'vehicle.'" They were then presented with ten sets of statements, in a random order. For example, a participant might first rate two statements appearing like this:

> An airplane is a prototypical vehicle. 1 (strongly disagree) to 7 (strongly agree)
>
> An airplane is technically a vehicle. 1 (strongly disagree) to 7 (strongly agree)

Results

As predicted, there were significant differences between the Prototypically and Technically judgments across the ten entities.[143] Comparing these results to Experiment 1's results for Corpus and Dictionary participants reveals a striking similarity.

---

[141] Eleanor Rosch, *Principles of Categorization*, *in* CONCEPTS: CORE READINGS 189, 197 (1999); *see also id.* at 189–205.

[142] *See* Appendices, app. B, pt. I, 134 HARV. L. REV. 726 app. (2020).

[143] *See infra* Figure 2.

Figure 2: Mean Ratings for "Prototypically" and "Technically" for Ten Entities[144]

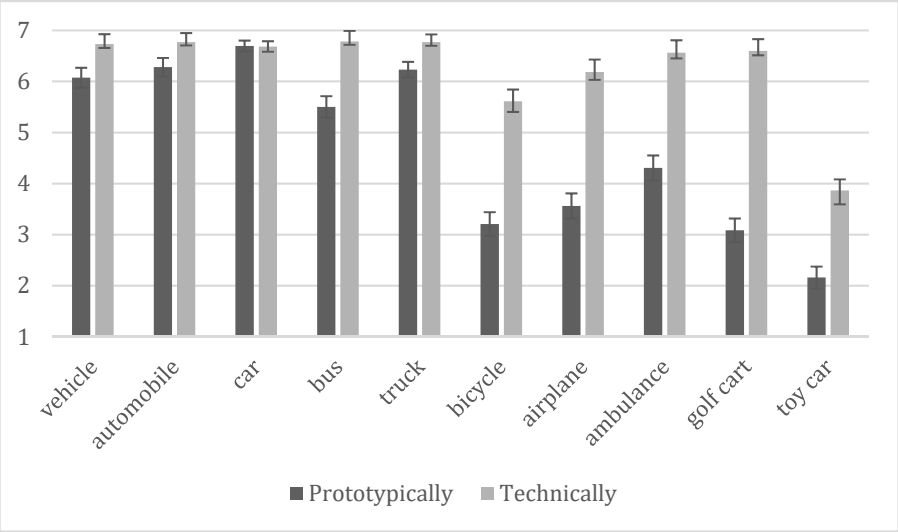
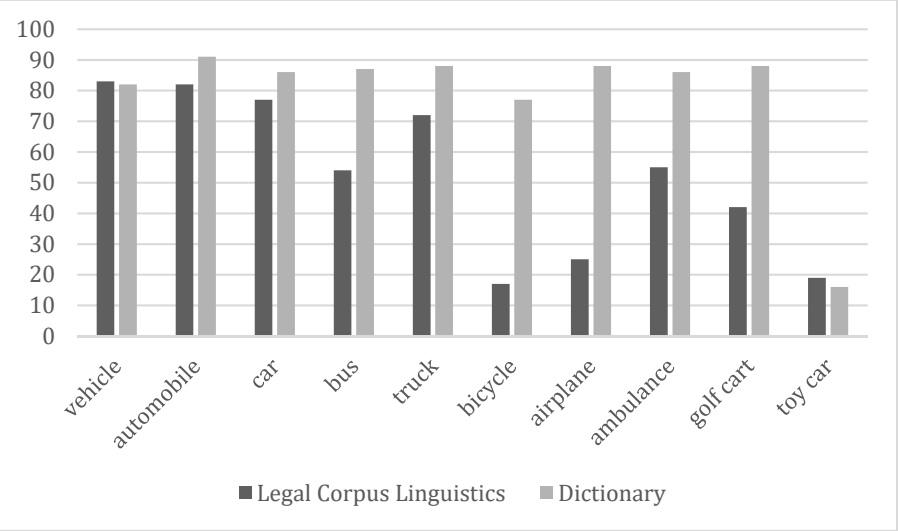


Figure 3: Percentage Responding "Yes" (Entity Is a Vehicle) for Ten Entities by Corpus and Dictionary Condition (Experiment 1)



144 Error bars indicate standard errors.

To test the statistical significance of this relationship between Corpus-Prototypically and Dictionary-Technically, I conducted two tests for differences in correlations between (i) Corpus and Prototypically and Corpus and Technically, and (ii) Dictionary and Prototypically and Dictionary and Technically.[145] In the first comparison, I considered the percentage of participants that rated each entity as a vehicle using the corpus, and correlated that percentage with the ratings for Prototypically and Technically. Prototypically was significantly more correlated with Corpus, $z = 1.841$, $p = 0.0328$ (one-tailed). Technically was significantly more correlated with Dictionary, $z = 3.49$, $p = 0.0002$ (one-tailed).[146]

Figure 4: Correlation Matrix

	Corpus	Dictionary	Concept	Prototypically	Technically
Corpus	1				
Dictionary	0.51	1			
Concept	0.73	0.93	1		
Prototypically	0.93	0.58	0.72	1	
Technically	0.72	0.95	0.99	0.74	1

Discussion

Insofar as legal corpus linguistics elicits more prototypical uses of a term but dictionaries elicit more extensive uses, the former may be more appropriate in legal contexts calling for a prototypical sense and the latter more appropriate in contexts calling for a more extensive sense. For example, in the context of a rule that "any and all vehicles are prohibited from the park," one might reasonably think that the ordinary meaning of the rule bans even bicycles. But in the context of a rule that "only cars, trucks, and other vehicles are prohibited from the park," one

---

145 *See* Ihno A. Lee & Kristopher J. Preacher, *Calculation for the Test of the Difference Between Two Dependent Correlations with One Variable in Common*, QUANTPSY.ORG (Sept. 2013), http://quantpsy.org/corrtest/corrtest2.htm [https://perma.cc/7YB2-9E2H].

146 Because the hypothesis was that Prototypically was more correlated with Corpus, and Technically with Dictionary, one-tailed tests were used rather than two-tailed tests. Two-tailed tests indicate similar results. Technically is significantly more correlated with Dictionary, $z = 3.489$, $p = 0.0005$ (two-tailed). Prototypically is more correlated with Corpus at a level short of the traditional cutoff for statistical significance, $z = 1.841$, $p = 0.0656$ (two-tailed). To match the previous experiments, this experiment uses only ten items — for example, car, bus, bicycle, and so forth. This limits the power of correlation analyses.

might argue more persuasively that the rule bans only more prototypical vehicles. If so, dictionaries would be better guides in the first case, but corpora would be better guides in the second. Part VI discusses these possibilities in greater detail.

C. Experiment 3: Expert Judges

The preceding experiments have studied populations with no expertise in law or interpretation. Judgments of ordinary people provide good evidence about the current ordinary meaning of these terms (for example, of "vehicle"). But some might doubt whether this population contains the best users of dictionaries and legal corpus linguistics. To appropriately test the reliability of legal corpus linguistics and dictionaries, one might argue, we should test legal experts who have the relevant background in interpretation.

This objection is plausible, but it should not be taken to dismiss any significance of the prior results. After all, even if the previous results do not provide strong evidence about judges' use of dictionaries and legal corpus linguistics data, they do provide good evidence about jurors' use of such sources. And jurors, too, are statutory interpreters.[147]

Nevertheless, this section addresses the "expertise" objection head-on. I tested a population composed of United States judges and law students from Harvard, Yale, and Columbia. There are a few reasons to think such experts might perform differently. For one, they may have some training or expertise that enables them to use dictionaries or legal corpus linguistics in some expert way. Additionally, even if they do not have special expertise in legal corpus linguistics per se, they might be more reliable survey takers, more likely to devote sufficient attention and produce thoughtful responses.

To more comprehensively test the reliability of dictionaries and legal corpus linguistics, this experiment featured an expanded range of twenty-five entities. In the first three experiments, most entities were "vehicles" in ordinary language and the dictionary categorized these as vehicles. The previous experiment suggested that the dictionary generates an extensive condition of category membership. So to better test dictionaries, this experiment also included some entities that I predicted are likely *not* vehicles in ordinary language, but which may nevertheless fall under a very extensive sense of a vehicle. These entities included crutches, a baby shoulder-carrier, and a zip line.

To avoid redundancy, I present only the judge data in the main text. The law student data (which is very similar) is presented in Appendix C.

---

147 Lawrence M. Solan, *Jurors as Statutory Interpreters*, 78 CHI.-KENT L. REV. 1281, 1282–83 (2003).

Method

Participants. Approximately seven hundred professional judges were contacted by email to request voluntary participation in the study. Ninety-six United States judges completed the online experiment.[148] Judges were recruited from state and federal courts and asked to categorize their years of experience. Seventy-four judges reported their years of judging experience. Of those, 1% reported less than one year of experience, 17.5% reported one to five years of experience, 19% reported six to ten years, 24% reported eleven to fifteen years, 17.5% reported sixteen to twenty years, 11% reported twenty-one to twenty-five years, and 8% reported over twenty-six years.[149] The same experiment was completed by two hundred one participants from MTurk (M_{age} = 37; 47% female, 53% male, 0% nonbinary) and two hundred thirty-two law students (M_{age} = 27; 48% female, 52% male, 0% nonbinary).

Materials and Procedure. As in the previous experiments, participants were randomly assigned to either the Concept, Legal Corpus Linguistics ("Corpus"), or Dictionary condition. In this experiment, participants evaluated the first set of entities (presented in a randomized order): a vehicle, automobile, car, bus, truck, bicycle, airplane, ambulance, golf cart, and toy car. Participants immediately considered another set (presented in a randomized order): a drone, skateboard, pair of roller skates, "a nonfunctioning commemorative truck (e.g. a World War II Truck that has been decorated as a World War II monument)," baby stroller, electric wheelchair, horse-drawn carriage, wooden canoe, helicopter, moped, pair of crutches, pogo stick, baby shoulder-carrier, life raft, and zip line. The binary categorization questions were of identical form to those posed in Experiment 1.

Results and Discussion

To analyze the data, I entered the binary categorization judgment as the dependent measure in a generalized linear mixed-effects model, with participant and entity as random effects, and Method (Corpus, Dictionary, Concept), Group (Lay, Law Student, Judge), and Method*Group fixed effects.[150]

That model revealed no significant Method*Group interaction effects, and a simpler model without the Method*Group term had no loss in explanatory power; X^2 = 2.24, p = .69; AIC of original model = 10,704; AIC of reduced model = 10,699.

In the final model, with Method and Group fixed effects and Participant and Entity random effects, there was a significant effect of

---

[148] M_{age} = 59.3; 34% female, 66% male, 0% nonbinary.

[149] 1% of respondents reported "other."

[150] *See* Stefan Th. Gries, *The Most Under-used Statistical Method in Corpus Linguistics: Multilevel (and Mixed-Effects) Models*, 10 CORPORA 95, 113 (2015).

Method (Corpus: OR = .04, 95% CI [.03, .06], z = -18.12, p < .0001; Concept: OR = .30, 95% CI [.22, .42], z = -7.36, p < .0001). There is also a significant effect of Group (Students: OR = .98, 95% CI [.73, 1.31], z = -1.29, p = .90; Judges: OR = 1.54, 95% CI [1.06, 2.25], z = 2.24, p = .025).

The results of the experts are strikingly similar to the results of the nonexperts from the earlier experiments. Like the lay participants, the law students' and judges' use of legal corpus linguistics and dictionary methods did not consistently reflect their ordinary judgments about category membership.

For many entities, the corpus linguistics judgment did not coincide with judges' ordinary conceptual competence: consider bus, truck, airplane, ambulance, golf cart, and horse-drawn carriage. For many others, the dictionary use did not reflect ordinary judgment: consider skateboard, roller skates, World War II Truck, baby stroller, canoe, helicopter, and baby shoulder-carrier.

For a very large number of entities, the corpus and dictionary delivered divergent judgments: truck, bicycle, airplane, golf cart, skateboard, roller skates, baby stroller, wheelchair, horse-drawn carriage, canoe, helicopter, baby shoulder-carrier, life raft, and zip line.

D. Experiment 4: Replication Across Ten Examples

Experiment 3 indicated that the main findings regarding dictionaries and legal corpus linguistics replicate across levels of legal expertise. Using the example of a vehicle, the experiment found that the verdicts delivered by dictionary use and legal corpus linguistics use often depart dramatically from each other and from the verdict delivered by ordinary judgment of language meaning.

This final experiment sought to test whether these findings replicate across different examples. To examine this question, the experiment tested "vehicle," as well as nine other terms. More broadly, the final experiment aimed to serve as a robustness check of the earlier findings. It altered various parameters from the first experiment: the relevant term, the corpus data used, and the dictionary definition used.

First, the experiment assessed ten terms. Of the ten, the first three were drawn from examples cited by corpus linguistics proponents: "vehicle," "carry," and "interpreter."[151] The next three were inspired by important interpretation terms: "labor," "tangible object," and "weapon" (a modern version of "arms"). The final four were common examples of large superordinate categories, which admit of a range of category members: "clothing," "furniture," "food," and "animal." For each term, the experiment asked about twenty-five entities.[152]

---

[151] *See* Lee & Mouritsen, *supra* note 14, at 836, 845, 848; *see also* Appendices, app. D, 134 HARV. L. REV. 726 app. (2020).

[152] For further detail, see Appendices, app. D, 134 HARV. L. REV. 726 app. (2020).

Moreover, while the earlier experiments used the News on the Web Corpus, this experiment used instead the Corpus of Contemporary English.

Finally, while the earlier experiments also used a representative dictionary definition cited by proponents of legal corpus linguistics (who are generally skeptical of dictionary use),[153] this experiment simply used the first full definition of the relevant term, from Merriam-Webster 2019 Online.[154] In some cases, these definitions supplied some examples alongside the definitions. For example, "vehicle" is defined as "a means of carrying or transporting something // planes, trains, and other vehicles : such as : a : motor vehicle, b : a piece of mechanized equipment."[155] This fourth experiment included two dictionary conditions. The first "full dictionary" condition included the entire first definition of the relevant term. The second "bare dictionary" condition included the definition, without examples. The "bare" definition for "vehicle" was "a means of carrying or transporting something."

The experiment also used a different fake term, "krob" rather than "ailac," for the corpus and dictionary conditions.

Method

Participants. I recruited 2,835 "general population" participants from the United States from Amazon's Mechanical Turk (M_{age} = 37.88; 46.1% female, 53.6% male, 0.3% nonbinary).

Materials and Procedure. The procedure was similar to that of the first three experiments. Participants were randomly assigned to one of four methods (ordinary concept, corpus, full dictionary, bare dictionary) and one of ten examples (vehicle, carry, interpreter, labor, tangible object, weapon, animal, clothing, food, furniture).[156]

Results and Discussion

To analyze the data I entered judgment as the dependent measure in a generalized linear mixed-effects model, with Participant, Entity, and Category as random effects, and Method as fixed effects. There was a significant effect of Method (Corpus: OR = .68, 95% CI [.58, .80], z = -4.66, p < .0001; Concept: OR = 2.05, 95% CI [1.76, 2.39], z = 9.20, p < .0001; Bare Dictionary: OR = 1.55, 95% CI [1.32, 1.83], z = 5.36, p < .0001). The results again indicate a large degree of divergence among the Concept, Corpus, "Full Dictionary," and "Bare Dictionary" conditions.

---

153 *See, e.g.*, Lee & Mouritsen, *supra* note 14, at 800.

154 *See generally* MERRIAM-WEBSTER, https://www.merriam-webster.com [https://perma.cc/VHG5-9T2A].

155 *Vehicle*, MERRIAM-WEBSTER, https://www.merriam-webster.com/dictionary/vehicle [https://perma.cc/E25Y-WKPW].

156 See Appendices, app. D, 134 HARV. L. REV. 726 app. (2020) for full materials.

E. Summary and Interpretation

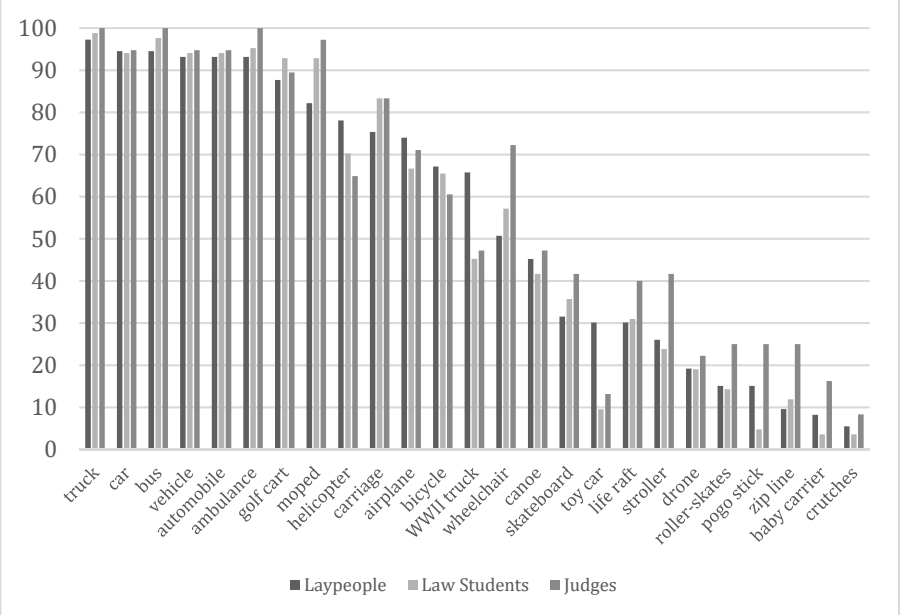
Before turning to the next Part, it is worth providing some summary considerations and graphics. The experiments suggest that judges and nonexperts are similar in (i) their ordinary judgments concerning common terms (for example, "vehicle"), (ii) how they apply dictionary definitions, and (iii) how they apply legal corpus linguistics data.

Moreover, the results indicate that, perhaps surprisingly, "ordinary meaning" is not as clear as one might think. For a number of entity categorizations, participants are very divided. For example, people are generally divided (about 50–50%) on whether a canoe is a vehicle. This is true across ordinary people, law students, and judges.

The pattern of results also indicates that dictionaries tend to be more inclusive than legal corpus linguistics. Legal corpus linguistics categorizations are correlated with judgments of prototypicality, while dictionary categorizations are more extensive. This implies that dictionaries and legal corpus linguistics often provide dramatically different verdicts from each other. Moreover, they often provide different recommendations about meaning from what is reflected in ordinary judgments.

1. Judges and Nonexperts Judge Meaning Similarly. — First, consider the percentage of participants within each population responding that each entity is a vehicle.

Figure 5: Percentage of Participants Responding "Yes" to "Is [Entity] a Vehicle?"



There is a striking similarity in the ordinary concept of a vehicle among those with very different legal and educational backgrounds. Whatever legal experience might provide, it does not seem to dramatically change cognition about ordinary concepts like vehicles.

It is worth noting that some responses — from all groups — seem obviously incorrect. For instance, about 5% of each population evaluated vehicles, cars, and buses as not "vehicles." This may have occurred for a number of reasons. Perhaps a small number of participants were answering randomly or incautiously. The fact that there is large variation among entities and consistent variation among groups suggests strongly that most participants were not answering in such a fashion. That is, if most of the lay participants answered randomly, we would expect percentages for all entities (from truck to crutches) to fall close to 50%. Instead, we see much variation across entities, and striking similarity in those judgments among lay, law-student, and judge samples.

Nevertheless, one might worry that a small number of unreliable survey takers (say, the 5% of lay participants who say a vehicle is not a vehicle) might threaten the interpretation of the data. It is a very difficult question how to translate the overall percentages (for example, 90% of people today evaluate a golf cart as a vehicle) into the "fact" or empirical construct of ordinary meaning (for example, today, the ordinary meaning of "vehicle" includes golf carts).

Section IV.E.9 considers various plausible ways of conducting this translation from the concept condition percentages into a verdict about ordinary meaning. For example, one plausible way to translate is to use a 50% cutoff: if 50% of people today categorize an entity as a vehicle, it is part of the ordinary meaning of "vehicle"; and if fewer than 50% categorize it as a vehicle, it is not part of the ordinary meaning. Other translations would use different cutoffs, such as a 75% cutoff. Still other translations make different kinds of assumptions.

Many of these translations accommodate some small number of mistakes or performance errors. For example, imagine using the 50% cutoff. If 5% of the law students mischaracterized a car as *not* a vehicle, this would not change the fact that, on a 50%-cutoff translation, a car is part of the ordinary meaning of "vehicle."

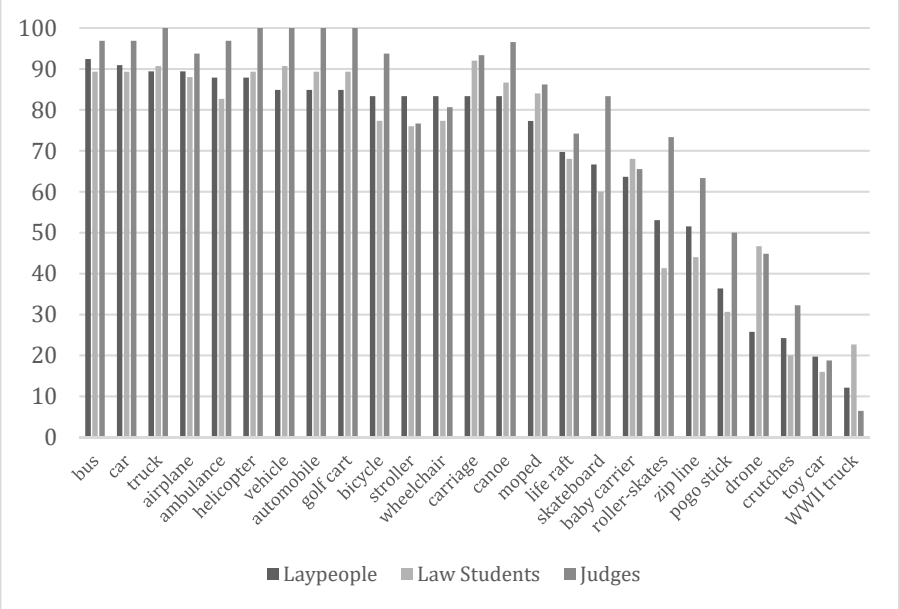
2. Ordinary Meaning Is Often Unclear or Indeterminate. — Figure 5 indicates a second striking fact. Although the results are similar among the three populations, there is notable disagreement among people about which entities are category members. For example, among judges, law students, and those untrained in law, there is substantial disagreement about whether canoes and skateboards are vehicles.

Interpreters typically seek to discover *a fact* about ordinary meaning, but this result suggests that in some cases such a fact may be unclear or indeterminate. Taken at face value, the results suggest that there is no clear fact of the matter concerning whether the modern ordinary meaning of "vehicle" includes a canoe. Moreover, this disagreement is not

mitigated by judicial or legal expertise. Disagreement persists to a similar degree across people with various levels of legal training and experience.

3. Judges and Nonexperts Apply Dictionaries Similarly. — Next consider the percentage of participants within each population responding that each entity is a vehicle, according to the dictionary.

Figure 6: Percentage of Participants Affirmatively Categorizing the Entity According to the Dictionary Definition

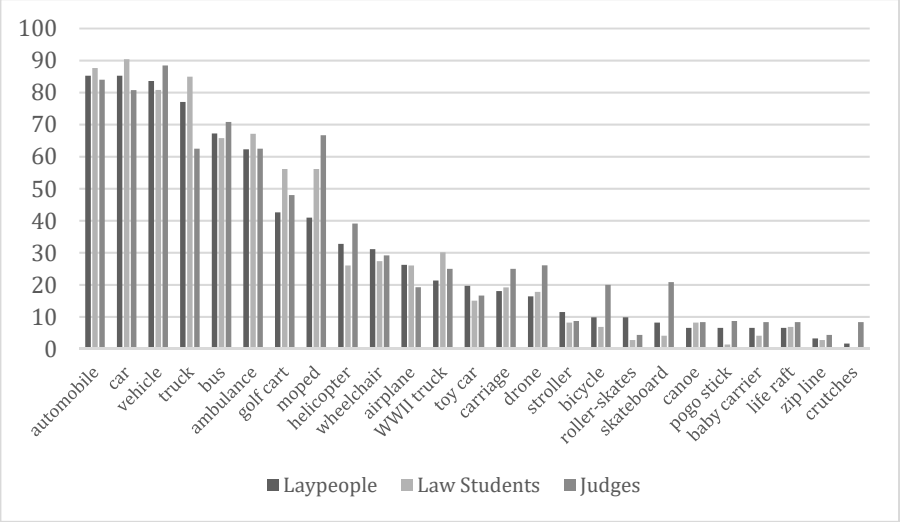


The results suggest that legal expertise does not dramatically change the way in which people apply a basic dictionary definition. Judges were modestly more inclined to categorize some entities as vehicles, but overall the pattern of results is fairly consistent among the three populations.

Again, it is also worth noting that although there is remarkable agreement among the populations — judges, law students, and lay participants do not disagree *as groups* about how to apply dictionaries — there is striking disagreement within groups for some entities. Consider examples like zip line, pogo stick, and drone. A substantial proportion of participants in every group categorized these as vehicles, while a substantial proportion did not.

4. Judges and Nonexperts Apply Legal Corpus Linguistics Similarly. — Finally, consider the percentage of participants within each population responding that each entity is a vehicle according to legal corpus linguistics.

Figure 7: Percentage of Participants Affirmatively Categorizing the Entity According to the Legal Corpus Linguistics Data

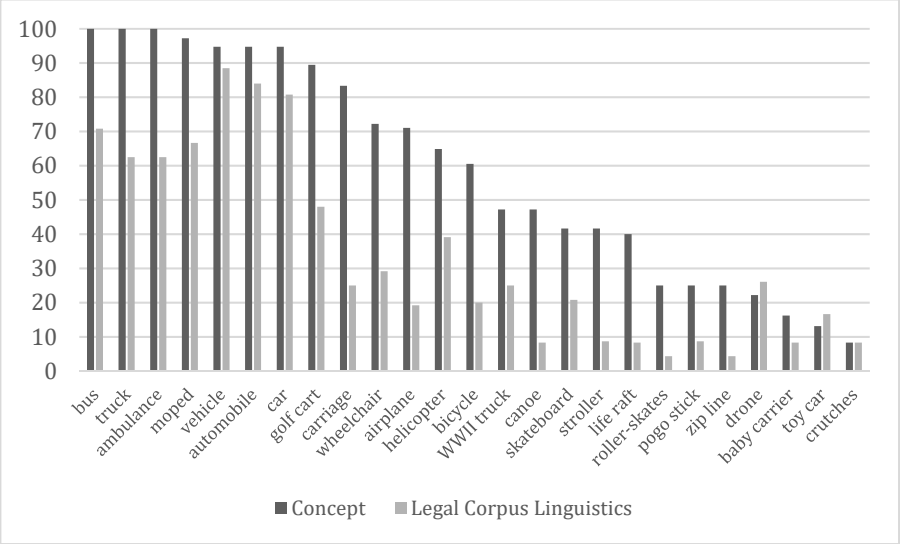


The results suggest that legal expertise does not dramatically change the way in which people interpret and apply the legal corpus linguistics data.

Here again, we should note that although there is impressive agreement among the populations — judges, law students, and lay participants are not very different *as groups* in their applications of corpus linguistics — there is striking disagreement within groups for some entities. Consider examples like the golf cart, moped, and helicopter. Within each group, a substantial proportion of participants categorized these as vehicles, while a substantial proportion did not.

5. Legal Corpus Linguistics Use Reflects Narrow, Prototypical Uses. — Next consider the corpus and concept results for judges.

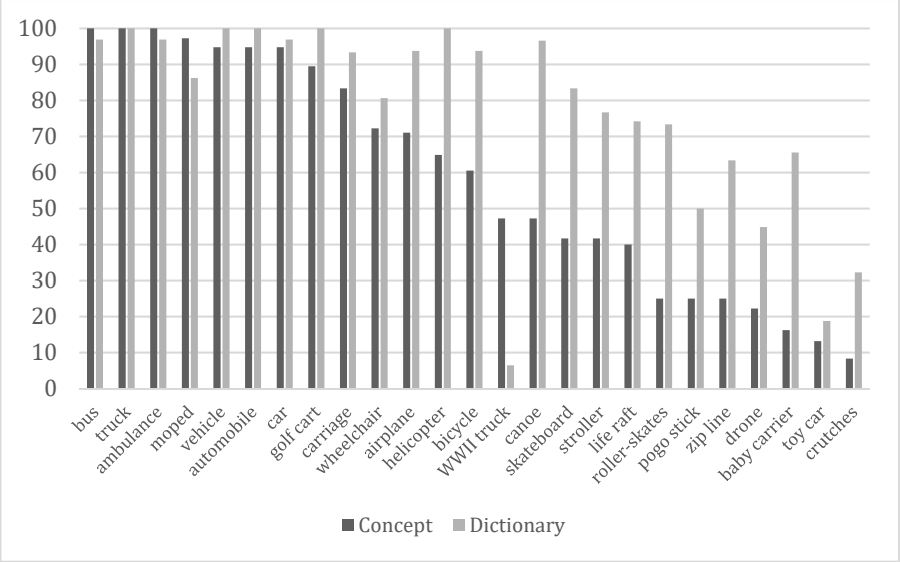
Figure 8: Comparison of Percentage of Judges Responding "Yes" to "Is [Entity] a Vehicle?" to Percentage of Judges Categorizing the Entity as a Vehicle According to Legal Corpus Linguistics



Although legal corpus linguistics use is not entirely unrelated to judges' application of the ordinary concept, in many cases the corpus is underinclusive. For example, a truck is unanimously understood as a vehicle in ordinary language, but users of legal corpus linguistics returned only a moderate endorsement of trucks as vehicles. Similarly, entities like horse-drawn carriages, golf carts, airplanes, helicopters, and bicycles are largely understood by judges as vehicles in ordinary language, but they are rarely classified as vehicles by judges using legal corpus linguistics.

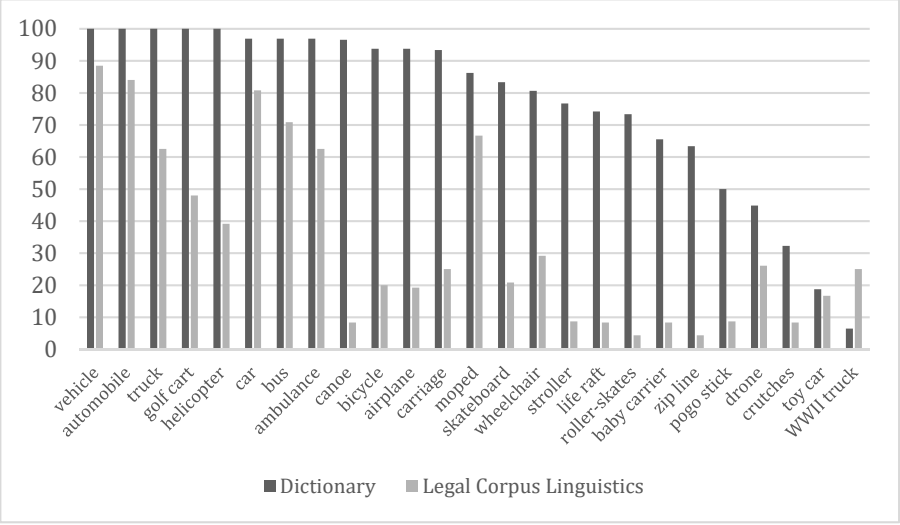
6. Dictionary Use Can Reflect Extensive Uses. — Next, consider the concept and dictionary results for judges. There are some large divergences between ordinary judgments and dictionary verdicts. For example, most using the dictionary evaluate baby-shoulder carriers as vehicles; however, we generally understand that those are not vehicles. For other controversial entities (for example, a canoe), dictionary use tends to indicate that those *are* vehicles.

Figure 9: Comparison of Percentage of Judges Responding "Yes" to "Is [Entity] a Vehicle?" to Percentage of Judges Categorizing the Entity as a Vehicle According to Dictionary Use



7. Dictionaries and Legal Corpus Linguistics Provide Divergent Verdicts. — It is worth considering the Dictionary and Corpus results together. These results reflect some dramatic differences between the results suggested by dictionaries and legal corpus linguistics. Insofar as these sources of evidence should be tracking the same "ordinary meaning," these results suggest some serious risk of error in at least one of the methods.

Figure 10: Comparison of Percentage of Judges Categorizing the Entity as a Vehicle According to Dictionary Use and Legal Corpus Linguistics Use



In many cases, these results reflect extreme differences between dictionaries and legal corpus linguistics. Nearly every judge using a dictionary assessed entities like canoes, bicycles, and airplanes as vehicles; while nearly every judge using legal corpus linguistics assessed those as *not* vehicles.

8. Dictionary and Legal Corpus Linguistics Verdicts Diverge from Ordinary Judgment. — Figures 9 and 10 indicate that, for many examples, definitions suggested by the use of dictionaries and legal corpus linguistics were very far from reflecting ordinary judgments.[157] The most straightforward interpretation of this pattern of results is that dictionaries and legal corpus linguistics were not always accurate measures of modern ordinary meaning.

Considering just the results of Experiment 3, in some cases, legal corpus linguistics indicated that seemingly clear vehicles were not, in

[157] *See also infra* apps. C–D.

fact, vehicles. For example, buses, trucks, and ambulances were unanimously understood to be vehicles. Yet over one-third of judges using legal corpus linguistics evaluated these as not vehicles.[158] Conversely, in some cases the dictionary use indicated that clear nonvehicles were, in fact, vehicles. For example, dictionary-using judges overrated roller skates and baby-shoulder carriers as vehicles, compared to judges' ordinary evaluation of those entities.

9. On Average, Legal Corpus Linguistics and Dictionaries Had 20–35% Error Rates. — This section computes and considers the "error rates" for both dictionary and legal corpus linguistics methods, across Experiments 3 and 4. That is, it considers how often one *relying solely* upon a dictionary definition or legal corpus linguistics data would reach the wrong verdict about ordinary meaning.

To assess these error rates, we must make some assumptions about what percentage of agreement in the ordinary concept condition indicates that the use is within the ordinary meaning. For example, should we assume that the ordinary meaning of "vehicle" includes a car if at least 50% of people agree; or must some higher threshold, like 75% or 90%, be met? Some scholars have suggested particular cutoffs, such as a "supermajority" of 60% or 67%.[159]

This Article takes no stand on this contentious issue. Rather, it considers three cutoffs, 50%, 75%, and 90%, as a representative range of plausible options. As such, this analysis does not require us to take a position on this hard question about ordinary meaning (we need not commit that 50% is the right cutoff to distinguish ordinary meaning from ordinary judgments). Instead, this analysis allows us to consider the error rates across a range of plausible options. As we will see, there is some similarity in the error estimated across these options. This allows us to conclude that, under many plausible assumptions, *relying solely* on a dictionary definition or corpus linguistics dataset would suggest the wrong verdict in a substantial number of cases.

To give a sense of how this computation works, consider a 50% cutoff. That is, assume that if over 50% of participants (in the ordinary concept condition) categorized something as a vehicle, then it is part of the ordinary meaning of "vehicle." To take one example, 100% of judges assessed a bus to be a vehicle. Because 100 is greater than 50, we treat this as a vehicle. Only 68% of judges using legal corpus linguistics made the same judgment. So, 32% of corpus users made a judgment (that is, that a bus is not a vehicle) that is incorrect given our assumptions. So given that assumption, there is a 32% error rate for the bus item for users of legal corpus linguistics in this judge sample. Repeating this process for all items (all 25 items in Experiment 3; and all 250 items in Experiment 4), we can compute an average error rate.

---

158 *See supra* section IV.C, pp. 762–64.

159 *Cf.* Ben-Shahar & Strahilevitz, *supra* note 71, at 1779; *see also id.* at 1780.

I performed these computations, using 50%, 75%, and 90% cutoffs, for the legal corpus linguistics and dictionary results from Experiment 3 (judges, law students, and general population evaluations of vehicles) and Experiment 4 (general population evaluations of ten examples). The results are displayed in the first three columns of Tables 2 and 3.

An alternative method of assessing error is to consider the "Difference Between Percents." On this method, we consider the absolute value of the difference between the percentages of affirmative judgments in the ordinary concept condition and one of the corpus or dictionary conditions. For example, for judges, 32.3% of dictionary users categorized crutches as a vehicle, and 8.3% of ordinary concept condition participants made the same categorization. So, the "Difference Between Percents" error rate for dictionaries for this item is 24.0% (32.3% minus 8.3%). As should be clear, this calculation of error is generous to legal corpus linguistics and dictionaries. The most natural interpretation of the crutches data is that it is *not* a vehicle in the ordinary sense; 8.3% of participants in the concept condition were wrong; and 32.3% of judges using dictionaries made an incorrect categorization. In this case, the "Difference Between Percents" method computes a dictionary error rate that is 8.3% lower. These results for corpus linguistics and dictionaries are displayed in the final column of Tables 2 and 3, respectively.

Table 2: Average Error for Legal Corpus Linguistic Judgments in Experiments 3 and 4, Under Different Theoretical Assumptions

Average Legal Corpus Linguistics Error				
	Ordinary Meaning = 50%	Ordinary Meaning = 75%	Ordinary Meaning = 90%	Difference Between Percents
Judges (Vehicle)	30.3%	22.8%	20.5%	22.4%
Law Students (Vehicle)	27.8%	21.0%	22.6%	25.0%
General Pop. (Vehicle)	32.5%	21.1%	18.4%	24.1%
General Pop. (Ten Examples)	41.9%	39.1%	39.5%	25.1%

Table 3: Average Error for Dictionary Judgments in Experiments 3 and 4, Under Different Theoretical Assumptions

Average Dictionary Error				
	Ordinary Meaning = 50%	Ordinary Meaning = 75%	Ordinary Meaning = 90%	Difference Between Percents
Judges (Vehicle)	29.9%	43.3%	50.8%	22.4%
Law Students (Vehicle)	28.6%	35.9%	46.0%	16.7%
General Pop. (Vehicle)	33.8%	41.9%	49.9%	21.7%
General Pop. (Ten, "Full")	34.2%	36.4%	41.3%	18.4%
General Pop. (Ten, "Bare")	35.1%	46.9%	47.6%	20.8%

The important takeaway from these tables is that the error rates for relying solely on a dictionary or legal corpus linguistics data are certainly not trivial. For judges, law students, and the general population (across many examples), the error rate for both tended to fall between 20% and 35%. Sometimes it was higher (for example, 50.8% for judges' use of dictionaries, with a 90% cutoff); and sometimes it was lower (for example, 18.4% for the general population in Experiment 4, with a 90% cutoff). But the results overwhelmingly indicate that these methods carry real risks of error. The range of error rates (20–35%) suggests that one

relying on dictionaries or legal corpus linguistics would reach the wrong verdict *once in every three to five cases*.

Importantly, this "error rate" is not an estimate of how often users of dictionaries or legal corpus linguistics reach the wrong verdict in actual practice. Some factors might lower that number, such as any interaction between the use of these methods and the user's understanding of the contractual or statutory context. However, there are a number of other factors, such as politically motivated reasoning, that might *increase* that number even further. The "error rate" represents something very different: it is the frequency of error we should expect *if an interpreter were to rely solely on the dictionary definition or legal corpus linguistics data concerning a term*. As such, the error rate calculation is most significant for that specific type of — not uncommon — legal interpretation.[160]

10. In Some Circumstances, Error Rates Reached 80–100%. — Although the notion of an average error rate is helpful, it is also useful to consider the maximum error rates. The experiments included a number of relatively easy categorizations (for example, whether a car is a vehicle; whether factory work is labor; whether a book is a tangible object). Insofar as real legal decisions concern comparatively more difficult categorizations (for example, whether an airplane is a vehicle; whether preaching is labor; whether a fish is a tangible object), it may also be instructive to consider the maximum error rate: What percent of judges, for example, using dictionary or corpus linguistics evaluated *the hardest interpretive question* incorrectly?

These experiments involved a similar analysis to that conducted in section IV.E.9. But this analysis computed the maximum error rate, under each of the different assumptions. As Tables 4 and 5 indicate, across all levels of expertise, the data suggest that in some examples, relying on a dictionary definition or legal corpus linguistics data led 80–100% of users to the incorrect verdict. For example, the 80.8% "Maximum Legal Corpus Linguistics Error" for judges (Table 4) at the 50% cutoff refers to the error rate regarding airplanes. The entry in the next column (75% at a 75% cutoff) refers to the error rate regarding horse-drawn carriages; the next entry (46.1% at a 90% cutoff) is for golf carts; and the final entry in that row (58.3% on a difference between percents calculation) refers to horse-drawn carriages.

---

[160] *See* Nourse, *supra* note 32, at 678–79.

Table 4: Maximum Error for Legal Corpus Linguistic Judgments in Experiments 3 and 4, Under Different Theoretical Assumptions

Maximum Legal Corpus Linguistics Error				
	Ordinary Meaning = 50%	Ordinary Meaning = 75%	Ordinary Meaning = 90%	Difference Between Percents
Judges (Vehicle)	80.8%	75.0%	46.1%	58.3%
Law Students (Vehicle)	92.5%	85.0%	90.0%	76.0%
General Pop. (Vehicle)	90.2%	67.2%	42.6%	57.7%
General Pop. (Ten examples)	75.3%	73.1%	73.5%	37.1%

Table 5: Maximum Error for Dictionary Judgments in Experiments 3 and 4, Under Different Theoretical Assumptions

Maximum Dictionary Error				
	Ordinary Meaning = 50%	Ordinary Meaning = 75%	Ordinary Meaning = 90%	Difference Between Percents
Judges (Vehicle)	96.6%	100%	100%	49.3%
Law Students (Vehicle)	87.2%	87.2%	94.4%	66.1%
General Pop. (Vehicle)	87.9%	89.4%	89.4%	57.7%
General Pop. (Ten, "Full")	85.1%	87.8%	93.9%	73.2%
General Pop. (Ten, "Bare")	86.7%	86.1%	88.6%	76.5%

These results indicate the potential gravity of the risk of error in relying on dictionaries and legal corpus linguistics in interpretation. In the hardest interpretive tasks, the percentages of judges, law students, and ordinary people reaching incorrect verdicts on the basis of legal corpus linguistics and dictionaries reached 50%, 75%, and even 100%.

V. How Jurists Use Dictionaries and Legal Corpus Linguistics

This Part considers to what extent the processes uncovered by the experiments manifest in real-world contexts: Do legal uses of dictionaries tend to reflect broad, extensive interpretations, while uses of legal corpus linguistics tend to reflect narrow, prototypical uses? While legal corpus linguistics is relatively new, judges frequently cite dictionaries.[161] Section V.A surveys the pattern of citation and finds that case law tends to refer to dictionary definitions as "broad" significantly more often than as "narrow." Moreover, while jurists often take dictionaries to support extensive senses of meaning, the definitions are sometimes narrowed by considering contextual features or which of multiple definitions is most relevant. Legal corpus linguistics has been used less frequently in case law, but many of the extant examples suggest that focusing on patterns of word usage leads to more narrow interpretations. Moreover, the dictionary-extensive, corpus-narrow relationship holds for several divisive examples. "Emoluments" seems broad when scholars survey Founding-era dictionaries, but narrow when scholars consider usage in historical corpora. So too for "commerce." The experimental insight about uses of dictionaries and legal corpus linguistics sheds light on these and other debates about ordinary meaning.

Although section V.A suggests that this tendency of dictionary and legal corpus linguistics methods manifests in legal decisionmaking, it is important to recall that there are many factors that might affect legal outcomes, and also many factors that might affect real-world uses of dictionary definitions and corpus linguistics. A very common and natural question is how the use of these tools interacts in the real world with political bias and motivated reasoning. There are too few legal corpus linguistics uses to adequately assess this claim, but section V.B considers whether the characteristics of dictionary definitions — they are generally broad but can be narrowed by some interpretive choices — may admit of political decisionmaking. To examine this question, the section considers two examples from the Bill of Rights, each of which contains three terms: the Second Amendment's "keep and bear Arms" and the Eighth Amendment's "cruel and unusual punishments."

Republican-appointed jurists, at the Supreme Court and circuit court levels, more frequently cite Founding-era dictionaries to interpret terms like "keep," "bear," and "arms" broadly. Conversely, when Republican-appointed jurists cite dictionaries in Eighth Amendment cases, they interpret the broad dictionary definitions narrowly. Although Democratic-appointed jurists cite dictionaries less frequently, when they

---

[161] *See* Recent Case, *supra* note 111, at 691 & n.6 (noting corpus linguistics first appeared in a federal judicial opinion in 2019); *see also* Samuel A. Thumma & Jeffrey L. Kirchmeier, *The Lexicon Has Become a Fortress: The United States Supreme Court's Use of Dictionaries*, 47 Buff. L. Rev. 227, 244–63 (1999) (analyzing in detail the Court's increased reliance on dictionaries).

do, the pattern is reversed: dictionaries support that "cruel and unusual punishment" is broad, but "keep and bear arms" is narrow.

Together, the two sections indicate that the experimental findings track an important aspect of real-world use of dictionaries and legal corpus linguistics, while there are also important limitations on the scope of that insight: where dictionaries and corpus linguistics are cited, there are a number of other factors (for example, a statute's precedent or purpose or extrinsic evidence of contracting parties' intentions) and biases (for example, politically motivated reasoning) that add further complexity.

A. Dictionaries Often Supply "Broad" Senses and Legal Corpus Linguistics "Narrow" Ones

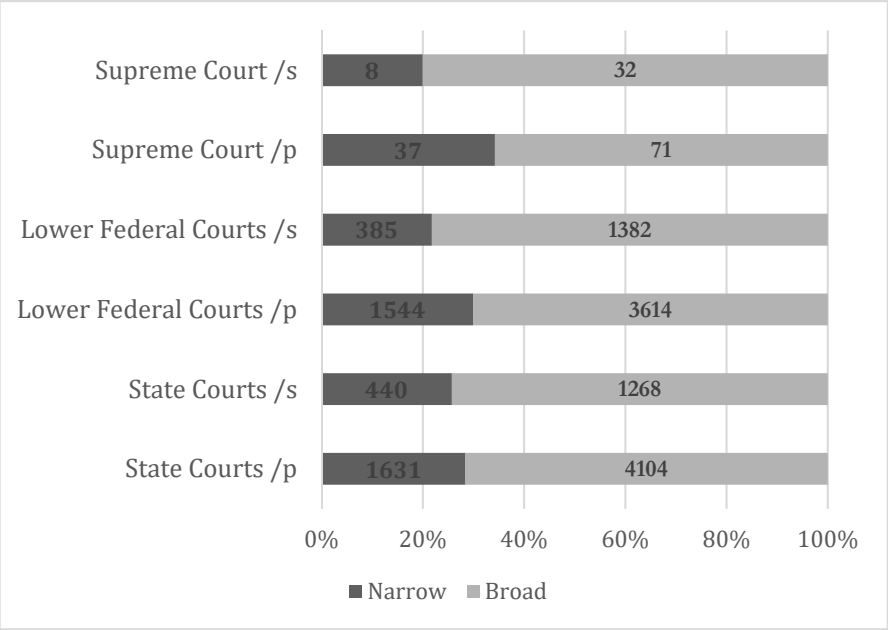
The experimental results indicate that for many (but not all) examples, dictionary definitions tended to reflect a broad sense of category membership. Those using the dictionary were inclined to include far more entities as category members, compared to those using legal corpus linguistics. And those using dictionaries were even inclined to categorize some entities as category members that are not judged to be category members in ordinary language. For example, dictionary users evaluated baby-shoulder carriers and roller skates as vehicles, but most people do not consider those entities to be vehicles.

This result may seem less surprising when reflecting on the nature of a dictionary. Dictionaries often present brief definitions that aim to comprehensively reflect a broad range of permissible uses. A "vehicle" is defined as a "means of carrying or transporting something." This definition is broader than what one might generate from considering only the most prototypical examples. For example, a car might be defined as an entity with four wheels that drives on roads. But using that as the definition of a vehicle would (inappropriately) exclude airplanes.

If this interpretation is right, we might expect courts' usage of dictionaries to reflect a similar sense that dictionaries provide broad definitions. As one approximation, consider courts' usage of the terms "broad" and "narrow" in the context of discussing dictionaries. Figure 11 shows uses of the terms "broad" and "narrow" within the same sentence or paragraph as the word "dictionary" in the Supreme Court, and a sample from lower federal courts and circuit courts.[162]

[162] For further detail, see *infra* notes 163–176175.

Figure 11: Court Citations of "Broad" and "Narrow" in the Same Sentence (/s) or Paragraph (/p) as "Dictionary"[163]



The data suggest that more often courts describe dictionary definitions as broad.[164] About 70% of the dictionary citations are near "broad" rather than "narrow." Compared to an estimation that citations would appear randomly — 50% near "broad" and 50% near "narrow" — this represents a statistically significant effect at all levels: for Supreme Court within-sentence uses,[165] Supreme Court within-paragraph uses,[166] lower federal court within-sentence uses,[167] lower federal court

---

[163] The data are drawn from searches conducted on Westlaw. "/s" indicates that the terms are within the same sentence; "/p" indicates that the terms are within the same paragraph.

[164] Some might wonder whether this corpus linguistics–style analysis can consistently be relied upon given the earlier critique of corpus linguistics. Importantly, this inquiry is very different from using corpus linguistics to establish public meaning. Corpus linguistics has a number of tremendously useful possibilities. The earlier critique is leveled at the claim that corpus linguistics reflects public meaning. That argument is independent from the claim that corpus linguistics provides evidence about whether dictionary definitions are typically described as broad or narrow.

[165] Binomial $p = .0099$.

[166] Binomial $p = .0275$.

[167] Binomial $p < .0001$.

within-paragraph uses,[168] state court within-sentence uses,[169] and state court within-paragraph uses.[170]

The same pattern of results holds true when taking into account the overall frequency with which courts use "broad" and "narrow." In one comparison, the effect is not statistically significant: Supreme Court within-paragraph uses.[171] However, for all other comparisons, the same pattern holds: for Supreme Court within-sentence uses,[172] lower federal court within-sentence uses,[173] lower federal court within-paragraph uses,[174] state court within-sentence uses,[175] and state court within-paragraph uses.[176] Overall, this pattern of results suggests that dictionaries are more often cited in the context of "broad" than "narrow."

Consider some of the Supreme Court's "broad" dictionary examples:

- "That a definition is broad enough to encompass one sense of a word does not establish, however, that the word is ordinarily understood in that sense."[177]
- "Just as the context of Rule 16 supports giving 'tangible object' a meaning as broad as its dictionary definition, the context of § 1519 tugs strongly in favor of a narrower reading."[178]
- "[T]he dictionary definitions of that word are very broad."[179]
- "Modern dictionaries contain . . . broad definitional language."[180]

One striking feature of these uses is that, while most suggest dictionary definitions are broad, many cite this breadth as a reason that legal interpretation should *not* follow the dictionary definition.

Of course, dictionaries are not uniformly understood to provide broad definitions. About 20–30% of the time, they are referred to near "narrow." In some of those examples, dictionaries are taken to provide a narrow definition:

---

168 Binomial $p < .0001$.

169 Binomial $p < .0001$.

170 Binomial $p < .0001$.

171 Binomial $p = .2478$.

172 Binomial $p = .0147$.

173 Binomial $p < .0001$. To provide a test case, these tests took the "broad" and "narrow" citations in the lower federal courts for three years, 2015, 2016, and 2017, and selected the highest ratio of broad/narrow uses, 63.9%. This selection provides a stringent test for the hypothesis.

174 Binomial $p < .0001$.

175 Binomial $p < .0001$. To provide a test case, these tests took the "broad" and "narrow" citations in state courts for January to June 2017 and July to December 2017 and selected the highest ratio of broad/narrow uses, 67.1.%. This selection provides a stringent test for the hypothesis.

176 Binomial $p < .0326$.

177 Taniguchi v. Kan Pac. Saipan, Ltd., 566 U.S. 560, 568 (2012).

178 Yates v. United States, 135 S. Ct. 1074, 1083 (2015) (quoting 18 U.S.C. § 1519).

179 Smith v. United States, 508 U.S. 223, 241 (1993) (Scalia, J., dissenting).

180 Bullock v. BankChampaign, N.A., 569 U.S. 267, 272 (2013).

- "To our knowledge *all* English dictionaries provided the narrow definition of 'modify' [connoting only moderate, and not fundamental, change] . . . ."[181]

However, many of the dictionary references near "narrow" actually suggest that dictionaries are understood to be broad:

- "Some [law and business dictionaries] define 'firm' [narrowly] . . . . But other dictionaries, while recognizing that narrow definition, also state that the word has a broader meaning . . . ."[182]
- "[C]ommon usage at the time of the National Bank Act prevents the conclusion that the Comptroller's refusal to give the word 'rate' the narrow meaning petitioner demands is unreasonable. The 1849 edition of Webster's gives as one of the definitions of 'rate' the '[p]rice or amount stated or fixed on any thing.'"[183]
- "When we have stated that sovereignty is a political question, we have referred not to sovereignty in the general, colloquial sense, meaning the exercise of dominion or power [citing the 1934 edition of *Webster's New International Dictionary*], but sovereignty in the narrow, legal sense of the term, meaning a claim of right."[184]

These empirical results are consistent with the experimental findings: dictionary definitions often (but not always) supply a broad, extensive sense of meaning. Importantly, however, they sometimes provide narrow meanings. Of course, there are some ways in which even broad dictionary definitions might be narrowed. Most notably, one might narrow a broad definition by (a) considering linguistic or legal context,[185] (b) selecting a relatively narrower definition, where there are multiple definitions,[186] or (c) emphasizing the necessity of particular features of the definition.

---

[181] MCI Telecomms. Corp. v. Am. Tel. & Tel. Co., 512 U.S. 218, 228 (1994).

[182] United States v. Cook, 384 U.S. 257, 261 n.5 (1966).

[183] Smiley v. Citibank (S.D.), N.A., 517 U.S. 735, 746 (1996) (second alteration in original) (quoting NOAH WEBSTER, AN AMERICAN DICTIONARY OF THE ENGLISH LANGUAGE 910 (1849)).

[184] Boumediene v. Bush, 553 U.S. 723, 754 (2008) (citation omitted).

[185] Compare, for example, Justice Breyer's and Justice Ginsburg's opinions in *Muscarello v. United States*, 524 U.S. 125 (1998), disputing whether the phrase "carries a firearm" should be read broadly because such a reading is consistent with common parlance and Congress's likely intent, *id.* at 128–31 (Breyer, J., majority opinion), or narrowly in keeping with the doctrine of lenity, *id.* at 148–49 (Ginsburg, J., dissenting).

[186] *See* Chisom v. Roemer, 501 U.S. 380, 410 (1991) (Scalia, J., dissenting) (arguing "that the ordinary meaning of 'representatives' does not include judges" (quoting 42 U.S.C § 1973(b) (current version at 52 U.S.C. § 10301(b))).

Legal corpus linguistics is relatively new and has yet to appear in a range of court decisions. So, inevitably, we can draw only more limited conclusions from case law practice. However, in the few cases that explicitly cite corpus linguistics, the results tend to narrow the contested sense of meaning.

For example, in the first opinion using corpus linguistics, *In re Adoption of Baby E.Z.*,[187] Justice Lee analyzed the phrase "custody determination."[188] He considered 500 sample sentences from the Corpus of Contemporary American English, and reported that the most common family law use of "custody" was in the context of divorce, rather than in the context of adoption.[189] He concluded that "the custody proceedings covered by the Act are limited to proceedings resulting in the modifiable custody orders of a divorce," rather than a broader range of custody proceedings.[190] This conclusion is consistent with the experimental findings, in which corpus linguistics often suggests that ordinary meaning is limited to prototypical uses.

Similarly, in *State v. Rasabout*,[191] corpus linguistics suggested that "discharge" was largely used to refer to a single shot of a firearm, rather than emptying the entire magazine.[192] This definition, too, is a narrower interpretation, limiting the ordinary meaning of "discharge" to the most common and prototypical use.

Finally, consider that in 2018 Justice Thomas made the first explicit reference to corpus linguistics in the Supreme Court. In a dissent regarding the meaning of "expectations of privacy," Justice Thomas noted that "[t]he phrase 'expectation(s) of privacy' does not appear in . . . the papers of prominent Founders, early congressional documents and debates, collections of early American English texts, or early American newspapers."[193] This dissent reflects a broadly similar use of corpus linguistics: the relative infrequency of a use from the corpus (in this case, the absence of a use) is taken to suggest that the use is not part of the original public meaning.

To be sure, evaluating linguistic usage data need not always provide a narrowing or exclusive recommendation concerning ordinary meaning. But the early judicial uses of corpus linguistics suggest such a trend.

This pattern of results — dictionaries tend to generate broader senses of meaning and legal corpus linguistics tends to generate narrower senses

---

187 J.M.W. v. T.I.Z. (*In re* Adoption of Baby E.Z.), 266 P.3d 702 (Utah 2011).

188 *Id.* at 719–25 (Lee, J., concurring in part and concurring in the judgment).

189 *Id.* at 724 & n.21.

190 *Id.* at 725.

191 356 P.3d 1258 (Utah 2015).

192 *Id.* at 1282 (Lee, A.C.J., concurring in part and concurring in the judgment).

193 Carpenter v. United States, 138 S. Ct. 2206, 2238–39 (2018) (Thomas, J., dissenting) (footnotes omitted) (citing corpus linguistics data).

of meaning, and the senses generated by each may be different — also helps explain some divisive debates about particular terms.

As one example, consider the recent debate about the original public meaning of "emolument" in the Constitution. One putative, "narrow" sense of the meaning is something like "profit arising from office or employ."[194] But another putative, "broad" sense need not involve "office" or "employ[ment]."[195] An impressive analysis of forty Founding-era dictionaries finds support for the broad interpretation.[196]

Conversely, a corpus linguistics analysis finds support for a less extensive meaning.[197] The study's authors report that the broad sense of "emolument" was more common than the narrow sense in an ordinary language corpus (20% more common), but the narrow sense was more common in "elite" and "legal" corpora (35% and 43% more common, respectively).[198] The paper concludes that the Congressional and Presidential Emoluments Clauses would have been understood to contain a narrow sense of "emolument," while the Foreign Emoluments Clause is more ambiguous.[199]

A similar debate arises in analysis, based on dictionaries and legal corpus linguistics, of the original public meaning of "commerce" in the Commerce Clause. Does "commerce" mean something broad like "any gainful activity" or "intercourse," or something narrower like "the trade and exchange of goods and transportation for this purpose"? Samuel Johnson's dictionary defines "commerce" broadly.[200] However, a thorough corpus linguistics–style examination of "every appearance of the word 'commerce' [in several Founding-era sources] . . . finds no surviving example of this term being used in this broader sense."[201]

These various examples indicate that the processes suggested by the experiments are consistent with real-world and scholarly practice. It is worth noting that jurists should not necessarily expect this to be the case, even if the experimental results are externally valid. The experiments are testing what dictionary definitions and corpus linguistics data

---

194 *See, e.g.*, *Emolument*, *in* JAMES BARCLAY, A COMPLETE AND UNIVERSAL ENGLISH DICTIONARY ON A NEW PLAN (1774).

195 *See id.*

196 John Mikhail, The Definition of "Emolument" in English Language and Legal Dictionaries, 1523–1806, at 8, 27 (July 12, 2017) (unpublished manuscript) (on file with the Harvard Law School Library).

197 James Cleith Phillips & Sara White, *The Meaning of the Three Emoluments Clauses in the U.S. Constitution: A Corpus Linguistic Analysis of American English from 1760–1799*, 59 S. TEX. L. REV. 181, 233 (2017).

198 *See id.* at 221.

199 *See id.* at 233–34.

200 Jack M. Balkin, *Commerce*, 109 MICH. L. REV. 1, 15 (2010) (citing SAMUEL JOHNSON, A DICTIONARY OF THE ENGLISH LANGUAGE (9th ed. 1790) (unpaginated)); *see also* Gibbons v. Ogden, 22 U.S. (9 Wheat.) 1, 1 (1824).

201 Randy E. Barnett, *The Original Meaning of the Commerce Clause*, 68 U. CHI. L. REV. 101, 101 (2001).

tend to suggest to interpreters — and not, for example, how judges typically weigh dictionaries against other sources of evidence or relevant interpretive factors. That is, even if dictionary use does in fact tend to reflect broad senses of meaning, while legal corpus linguistics use tends to reflect narrower senses, it could be that in actual legal practice, such initial reflections are outweighed or overruled — by licit factors including a statute's precedent or purpose or extrinsic evidence of contracting parties' intentions, or illicit factors such as politically motivated reasoning. This issue is generally outside the scope of this Article. Nevertheless, given the common interest in such questions, the next section considers the role of politically motivated reasoning.

B. Political Uses of Dictionaries and Corpus Linguistics

If dictionaries often provide broad, extensive senses of meaning, we should expect that jurists who cite dictionaries *should* reach inclusive or exclusive conclusions when dictionaries are cited at equal rates across similar types of cases. In other words, if dictionaries often reflect a broader sense of a term (say, 70% or 80% of the time), we might expect that citations of dictionary definitions lead to an inclusive interpretation at similar rates (for example, 70% or 80% of the time).

However, one might also wonder whether jurists sometimes use dictionaries politically.[202] If so, we might expect that jurists who cite dictionaries reach narrow or exclusive interpretations of the definition at surprisingly high rates when that narrow interpretation is consistent with the outcome associated with their political affiliation.

As one example, consider the contrast between two important clauses from the Bill of Rights: the right to "keep and bear Arms"[203] and the protection against "cruel and unusual punishments."[204] Broadly speaking, modern Republicans would prefer the former right interpreted broadly and the latter protection narrowly, while modern Democrats would prefer the former right interpreted narrowly and the latter protection broadly. But what do judges actually do?

Consider how Republican-appointed and Democratic-appointed federal jurists interpret dictionaries to support broad interpretations in Second and Eighth Amendment cases. First take "keep and bear arms." The only case in which the Supreme Court has used dictionaries to interpret these Second Amendment terms is *District of Columbia v. Heller*.[205]

---

[202] *See, e.g.*, James J. Brudney & Lawrence Baum, *Dictionaries 2.0: Exploring the Gap Between the Supreme Court and Courts of Appeals*, 125 YALE L.J.F. 104, 119 (2015) (considering whether the Supreme Court may rely on dictionaries partly due to its political visibility and resulting desire to appear neutral).

[203] U.S. CONST. amend. II.

[204] *Id.* amend. VIII.

[205] 554 U.S. 570 (2008).

But this is a rich case. The majority cites dictionaries to interpret all three terms, "keep," "bear," and "arms."[206] And the dissent also cites a dictionary to interpret "bear arms."[207]

Begin with the majority holding, authored by Republican-appointed Justice Scalia. He cites dictionary definitions of "keep," "bear," and "arms."[208] For "arms," Justice Scalia cites Samuel Johnson's 1773 dictionary, which "defined 'arms' as '[w]eapons of offence, or armour of defence.'"[209] He also cites Timothy Cunningham's 1771 legal dictionary, which "defined 'arms' as 'any thing that a man wears for his defence, or takes into his hands, or useth in wrath to cast at or strike another.'"[210] Justice Scalia also cites, but does not print, Noah Webster's 1828 definition.[211]

In addition, Justice Scalia cites dictionary definitions of "keep." He cites Johnson for the claim that "keep" meant "most relevantly, '[t]o retain; not to lose,' and '[t]o have in custody.'"[212] Moreover, "Webster defined it as '[t]o hold; to retain in one's power or possession.'"[213] Thus, Justice Scalia concludes, "the most natural reading of 'keep Arms' in the Second Amendment is to 'have weapons.'"[214]

Finally, Justice Scalia cites Johnson, Webster, Sheridan, and the Oxford English Dictionary for the claim that "bear" meant "carry."[215] For all three terms, the dictionary definition is understood to convey a broad sense of meaning, one that is inclusive in the context of *Heller*.

Conversely, the *Heller* dissent, which both Democratic-appointed Justices signed, cites "bear arms" as an idiom, finding that its dictionary meaning is "to serve as a soldier, do military service, fight."[216] It also cites the very same Johnson dictionary definition that Justice Scalia cites — "weapons of offence, or armour of defence" — but understands it to apply narrowly, exclusive of the use contested in *Heller*.[217]

---

206 *See id.* at 581–84.

207 *Id.* at 646–47 (Stevens, J., dissenting).

208 *Id.* at 581–84 (majority opinion).

209 *Id.* at 581 (alteration in original) (quoting *Arms*, 1 SAMUEL JOHNSON, A DICTIONARY OF THE ENGLISH LANGUAGE (4th ed. 1773) (reprinted 1978)).

210 *Id.* (quoting *Arms*, 1 TIMOTHY CUNNINGHAM, A NEW AND COMPLETE LAW-DICTIONARY (2d ed. 1771)).

211 *Id.* (citing *Arms*, 1 NOAH WEBSTER, AN AMERICAN DICTIONARY OF THE ENGLISH LANGUAGE (Found. for Am. Christian Educ. eds. 1989) (1828) (defining "arms" as "[w]eapons of offense or armor for defense and protection of the body")).

212 *Id.* at 582 (alterations in original) (quoting *Keep*, JOHNSON, *supra* note 209).

213 *Id.* (alteration in original) (quoting *Keep*, WEBSTER, *supra* note 211).

214 *Id.*

215 *Id.* at 584.

216 *Id.* at 646 (Stevens, J., dissenting) (quoting *Bear Arms*, 1 OXFORD ENGLISH DICTIONARY (2d ed. 1989)).

217 *Id.* at 647 (quoting *Bear Arms*, 1 SAMUEL JOHNSON, A DICTIONARY OF THE ENGLISH LANGUAGE (3d ed. 1755)).

Contrast this with the use of dictionaries in Eighth Amendment cases. In *Furman v. Georgia*,[218] Justice White refers to the broad dictionary sense of "cruel": "The imposition and execution of the death penalty are obviously cruel in the dictionary sense."[219]

However, more recent conservative-authored opinions use dictionaries to construe the Eighth Amendment's protection narrowly. Consider Justice Thomas's argument in *Baze v. Rees*[220] that lethal injections for executions are constitutional:

> Embellishments upon the death penalty designed to inflict pain for pain's sake also would have fallen comfortably within the ordinary meaning of the word "cruel." See 1 S. Johnson, A Dictionary of the English Language 459 (1773) (defining "cruel" to mean "[p]leased with hurting others; inhuman; hard-hearted; void of pity; wanting compassion; savage; barbarous; unrelenting"); 1 N. Webster, An American Dictionary of the English Language 52 (1828) (defining "cruel" as "[d]isposed to give pain to others, in body or mind; willing or pleased to torment, vex or afflict; inhuman; destitute of pity, compassion or kindness").[221]

It is worth considering the full definition of "cruel" cited in these dictionaries.

First, take "cruel" in Johnson's 1773 dictionary:

1. Pleased with hurting others; inhuman; hard hearted; barbarous. *Dryden*.
2. [Of things.] Bloody; mischievous; destructive. *Psalms*.[222]

Now consider "cruel" in Webster's 1828 dictionary:

1. Disposed to give pain to others, in body or mind; willing or pleased to torment, vex or afflict; inhuman; destitute of pity, compassion or kindness; fierce; ferocious; savage; barbarous; hardhearted; *applied to persons or their dispositions*.

   They are *cruel*, and have no mercy. [*Jeremiah* 6:23].

2. Inhuman; barbarous; savage; causing pain, grief or distress; exerted in tormenting, vexing or afflicting.

   Cursed be their wrath, for it was *cruel*. [*Genesis* 44:1].
   The tender mercies of the wicked are *cruel*. [*Proverbs* 12:10].
   Others had trials of *cruel* mockings. [*Hebrews* 11:36].[223]

---

[218] 408 U.S. 238 (1972).
[219] *Id.* at 312 (White, J., concurring).
[220] 553 U.S. 35 (2008).
[221] *Id.* at 97 (Thomas, J., concurring in the judgment) (alterations in original).
[222] *Cruel*, JOHNSON, *supra* note 209.
[223] *Cruel*, WEBSTER, *supra* note 211.

It is striking that Justice Thomas uses the definitions that are applied to persons, rather than the definitions applied to "*things*" (like the Eighth Amendment's "punishment"). Although the definitions relevant to persons appear first, it would seem that the definitions relevant to punishment (a thing) may be more apt. Understanding "cruel" punishments as ones that are "destructive" or "causing pain, grief or distress" suggests a much broader ordinary meaning.

Republican-appointed Supreme Court Justices apply dictionary definitions similarly (exclusively or narrowly) when defining "unusual." For example, in *Harmelin v. Michigan*,[224] the Court considered whether the imposition of mandatory sentences of life in prison without the possibility of parole, and without any consideration of mitigating factors, constituted cruel and unusual punishment.[225] Justice Scalia wrote for the majority, concluding that such punishment was not "unusual."[226] Unusual meant, according to Justice Scalia, "'such as [does not] occu[r] in ordinary practice,' Webster's American Dictionary (1828), '[s]uch as is [not] in common use,' Webster's Second International Dictionary 2807 (1954)."[227] Because mandatory sentences had been imposed regularly throughout history, they were not "unusual" in the Eighth Amendment context.[228]

Finally, in *Farmer v. Brennan*[229] and *Helling v. McKinney*,[230] the conservative Republican-appointed Justices' opinions indicate a narrow dictionary construal of "punishment." To them, punishment does not include an attack on a prisoner,[231] and punishment, not jail conditions, is the penalty for the commission of a crime.[232]

This pattern is consistent with the limited evidence from circuit courts.[233] Of all cases citing dictionaries in the same sentence as "cruel," "unusual," or "punishment," only one of those defined an Eighth Amendment term. This is *Duckworth v. Franzen*,[234] in which Judge Posner cited Johnson's dictionary to support the proposition that

---

[224] 501 U.S. 957 (1991).

[225] *Id.* at 994–96.

[226] *Id.*

[227] *Id.* at 976 (alterations in original).

[228] *Id.* at 994–95.

[229] 511 U.S. 825 (1994).

[230] 509 U.S. 25 (1993).

[231] *See id.* at 38 (Thomas, J., dissenting).

[232] *See Farmer*, 511 U.S. at 837–38; *id.* at 859 (Thomas, J., concurring in the judgment). *But see id.* at 854–55 (Blackmun, J., concurring) (citing dictionary definitions of "punishment" to argue against the other Republican-appointed Justices' "unduly narrow" interpretation, *id.* at 855).

[233] I considered all Westlaw-listed cases heard in federal courts of appeals that cite the Eighth Amendment. I searched within those for uses of "dictionary" in the same sentence as "cruel," "unusual," or "punishment." Nine cases were returned, one of which used a dictionary to define an Eighth Amendment term.

[234] 780 F.2d 645 (7th Cir. 1985).

"punishment" does not include injuries sustained when a bus to which prisoners were chained caught fire.[235] He argued that the dictionary definition requires that punishment be deliberate or reckless in the criminal law sense.[236] A broader pattern emerges when dictionaries are used by courts to decide whether legal texts are exclusive (narrow construal) or inclusive (broad construal). Recall Figure 11, which suggests that across all levels of the judiciary, dictionaries tend to admit of "broad" interpretations about 70% of the time and "narrow" interpretations about 30% of the time.[237] *If* we expect broad interpretations to imply inclusive legal determinations and narrow interpretations to imply exclusive legal determinations, we should find similar proportions across issues and political ideologies.

However, what we have found in Second and Eighth Amendment case law at the Supreme Court does not reflect this pattern. Instead, Republican appointees tend to construe dictionary definitions broadly when interpreting the terms "keep," "bear," and "arms," but narrowly when interpreting the terms "cruel," "unusual," and "punishment." Democratic appointees use dictionaries much less often, but when they do the pattern reverses: dictionaries indicate that Second Amendment terms are narrow, but Eighth Amendment terms are broad.

To be sure, this section has considered a very small sample. Future work may provide further insight into the question of whether and how dictionaries are used politically. The modest empirical analysis here suggests that it is a worthwhile question. In contrast, given the novelty of legal corpus linguistics, it is difficult to assess its political uses. However, careful interpreters may be wise to keep watch of emerging patterns.

VI. IMPLICATIONS

This Part turns to the experimental findings' implications for the theory and practice of legal interpretation. Section A elaborates on how the experiments clarify one of the processes underlying reliance on dictionary definitions or legal corpus linguistics data. Specifically, dictionary definitions tend to suggest broad senses of category membership, while word usage data tend to suggest more narrow, prototypical senses of category membership.

Section B identifies several fallacies often arising with the use of legal corpus linguistics and dictionaries that are supported by the experimental data. For example, consider "The Nonappearance Fallacy," the mistaken assumption that the nonappearance of some use in a corpus indicates that this use is outside of ordinary meaning. Arguments committing this fallacy have great rhetorical strength: *across thousands of*

---

[235] *See id.* at 652.
[236] *See id.* at 652–53.
[237] *See supra* Figure 11 & pp. 780–81; *see also supra* pp. 781–84 (explaining these findings).

sources in our corpus, we could not find even one example of an airplane referred to as a "vehicle"; therefore the ordinary meaning of "vehicle" does not include airplanes. However, as the experimental results indicate,[238] ordinary meaning sometimes diverges from ordinary use: people's understanding of language is not always reflected in recorded speech and writing, especially their understanding concerning nonprototypical category membership.

Section C considers implications for a set of interpretive theories that rely heavily on dictionary definitions or legal corpus linguistics to determine legal outcomes. This includes certain formalist, textualist, and originalist views on which a dictionary definition or sets of legal corpus linguistics data might be treated as sufficient to determine "the ordinary meaning" of a text and thereby determine the legal outcome. This section develops a broader burden-shifting argument. The experiments provide evidence that relying solely on dictionaries or legal corpus linguistics in determining ordinary meaning leads to significant and systematic errors — divergences between the methods and divergences from actual people's understanding of the relevant terms and phrases. Given the experimental results, interpretive theories relying on these methods have the argumentative burden of elaborating a nonarbitrary and demonstrably reliable use of legal corpus linguistics and dictionaries in interpretation.

Finally, section D evaluates the experimental results from the perspective of interpretive theories that are uncommitted to, or even skeptical of, the notion of a single "ordinary meaning" that determines legal outcomes across a range of cases and contexts. On these views, the experimental findings illuminate two different criteria that are often relevant in assessing the meaning of legal texts: a more extensive criterion and a more narrow, prototypical criterion. In many circumstances, dictionaries and legal corpus linguistics will help us assess each of these criteria, but a hard legal-philosophical question remains: Which of these two criteria *should* guide the interpretation of terms and phrases in legal texts? Insofar as there are good reasons underlying both criteria, the results suggest that dictionary definitions, legal corpus linguistics, or even other more scientific measures of meaning may not be equipped *in principle* to deliver simple and unequivocal answers to inquiries about the "ordinary meaning" of terms and phrases in legal texts.

A. Understanding the Use of Dictionaries and Corpora in Interpretation

Recall the experimental results that shed light on the psychological processes underlying use of dictionaries and legal corpus linguistics in

---

[238] *See supra* Part IV, pp. 753–77.

interpretation.[239] Verdicts from dictionaries were more strongly correlated with a term's extensivist uses than its prototypical ones. And verdicts from legal corpus linguistics were more strongly correlated with a term's prototypical uses than its extensivist ones.

For example, consider that a car is a more prototypical vehicle but an airplane is not.[240] Most participants using the dictionary were inclined to classify both entities as vehicles, but most participants using the corpus data classified only the (more prototypical) car as a vehicle.[241] Similarly, when participants considered the meaning of "carrying" a firearm, those using the dictionary provided more extensive judgments. Those using corpus data largely categorized prototypical examples as carrying (for example, taking a gun to a gang fight), but more often excluded less prototypical examples (for example, driving to a drug deal with a gun in the rear of the car).[242]

These results suggest something about the cognitive mechanisms underlying the use of these tools. While dictionary definitions help identify more extensivist uses of the term, legal corpus linguistics data tend to help identify more prototypical uses. There is something initially puzzling about this finding: legal corpus linguistics data are far more extensive than a brief dictionary definition, yet it is the latter that reflects a more extensive sense of meaning. Somewhat counterintuitively, copious corpus data produce relatively narrow judgments about meaning.

But this puzzle dissipates upon reflection. The standard dictionary provides definitions for *many* words, in a relatively compact space; to achieve this task, it is sensible to provide brief definitions that encompass broader senses of meaning. Conversely, legal corpus linguistics as typically practiced — focused on frequency analysis — identifies the most *common uses* of a term or phrase. It is unsurprising that this will underrepresent or omit very unusual uses, but (perhaps more surprisingly) it also underrepresents or omits even nonprototypical ones, such as the use of "vehicle" to describe an airplane.[243]

An important objection may be raised here: Although this pattern holds for the examples in this paper, should we infer that this reflects a broader pattern of judgment for many terms and phrases? To answer this challenge, first recall that the examples here were not chosen arbitrarily. The first two examples — vehicles and carrying a firearm — are

---

[239] *See supra* section IV.B, pp. 758–62.

[240] *See supra* section IV.B, pp. 758–62.

[241] *See supra* section IV.B, pp. 758–62..

[242] *See* Appendices, app. A, 134 HARV. L. REV. 726 app. (2020).

[243] *See also* Lawrence M. Solan, *Corpus Linguistics as a Method of Legal Interpretation: Some Progress, Some Questions*, 33 INT'L J. SEMIOTICS L. 283, 290 (2020) (explaining the "blue pitta problem").

two historically famous cases of statutory interpretation problems.[244] They were chosen for this reason and because they are the first two examples used by Lee and Mouritsen's recent manifesto on legal corpus linguistics.[245] That is, these are examples selected by other researchers who were unaware of the present hypotheses. Moreover, Lee and Mouritsen's paper is the leading *defense* of legal corpus linguistics; to the extent that these results cast doubt on legal corpus linguistics, they do so with regards to cases selected by its proponents.[246]

But maybe the original paper from which these examples are drawn just happened to select two unusual examples. That remains an open empirical question. Given the findings of this Article, I would expect the legal corpus linguistics and dictionary divergence to be widest when the investigated term is a category admitting of prototypical and non-prototypical members, and more narrow when that is not the case.[247] Additional data on other examples will enrich our understanding of dictionaries and legal corpus linguistics further.

There are several reasons to expect that this pattern of results would extend to other terms that admit of prototypical and more extensive uses. First, consider the type of evidence supplied by legal corpus linguistics. Advocates of legal corpus linguistics recommend data from collocation and "keywords in context" (concordance line) searches.[248] Collocation searches reflect the words that appear most frequently with the relevant search term.[249] There is good reason to think that many of these most common collocates are also representative of the features that we commonly attribute to the entity. For example, "vehicle" often appears near "electric," "gas," and "motor." We might also represent those as core features associated with vehicles. According to prototype theory, more prototypical members are the ones with more of the central features. So, insofar as the statistically common collocates also reflect our core associations with the concept, there is good reason to think that collocation is especially useful in identifying prototypical category members.

---

[244] *See generally, e.g.*, Lee & Mouritsen, *supra* note 14; Solan & Gales, *supra* note 44.

[245] *See* Lee & Mouritsen, *supra* note 14, at 800.

[246] *See generally id.*

[247] However, even well-defined terms like "even," "odd," "female," and "plane geometry figure" elicit judgments similar to judgments of prototype concepts. For example, the number four has been judged a better example of an even number than thirty-four; mother is a better example of a female than an actress; and a circle is a better example of a plane geometry figure than an ellipse. *See* Sharon Lee Armstrong, Lila R. Gleitman & Henry Gleitman, *What Some Concepts Might Not Be*, 13 COGNITION 265, 276 tbl.1 (1983).

[248] Lee & Mouritsen, *supra* note 14, at 831–32 (describing the potential value of using data from collocation to describe how words are most frequently used in context).

[249] *Id.*

The second type of legal corpus linguistics data, keywords in context, might also be especially useful in identifying prototypical category members. Keywords in context searches return example sentences from the corpus. We might think that such a search would return many types of uses — prototypical and nonprototypical. However, in practice, pragmatic considerations might limit the number of nonprototypical uses that we find.[250]

Consider these example sentences:

(1) Did you see any fish in the ocean?
(2) Look at that bird!
(3) The painter will finish painting Mike's fence tomorrow.

We might expect to find a sentence like (1) that refers to prototypical fish like trout or carp. It would not only be uncommon, but also seemingly *inappropriate* to say (1) if we meant to refer to sharks — instead we would ask "did you see any sharks in the ocean?" It would be even more unusual to say (1) if we meant to refer to stingrays. Of course, this pragmatic consideration in no way undermines the fact that people understand that sharks and stingrays are fish.

Similarly, (2) might occur when someone describes a prototypical bird like a robin or sparrow. We would expect to see examples like that in our modern corpus. But it would be a strange way to call attention to a penguin — even though penguins are birds. And we would probably not find many of these kinds of examples referring to penguins in the corpus.

In the same way, we would expect to find a sentence like (3) that refers to a prototypical painter (for example, an adult who works as a painter). Of course, if Mike's twelve-year-old niece enjoys painting and will paint his house, (3) could refer to her. But it would be strange, even inappropriate, to say (3) in that context. Instead, we would probably say something like, "Mike's niece will finish painting Mike's fence tomorrow." Mike's niece is still a painter, and anyone familiar with that fact would agree that she is a "painter" in the ordinary sense of the term in this context. Nevertheless, pragmatically, we would not usually say something like (3) if we meant to convey that Mike's niece will paint.

Now consider some of our legal examples:

(4) Asaf said we have to renew the vehicle registration.

A similar phenomenon operates here. It is possible that (4) could refer to an airplane registration, but it is more likely that we take the sentence to indicate a car registration. Using (4) would be a strange, if not inappropriate, way to describe an airplane. Consider a final example:

---

[250] *See generally* H. Paul Grice, *Logic and Conversation*, *in* 3 SYNTAX AND SEMANTICS 41 (Peter Cole & Jerry L. Morgan eds., 1975).

(5) Jasmin carried her books to school.

This would be an appropriate way to express that Jasmin hand-carried books to school. Of course, it could also express the fact that Jasmin loaded books into a wheelbarrow and towed them to school. But to express that, we would probably say something more specific than (5). Nevertheless, it is still true (in the ordinary sense of the term "carry") that Jasmin carried the books to school.

These examples suggest an intriguing phenomenon. Often, it is pragmatically inappropriate to refer to nonprototypical-category members by the broader category description. If we want to point out sharks in the water, we don't say "look at those big *fish*!"[251] This pattern of usage is perfectly consistent with the fact that sharks are understood to be fish (they are part of the ordinary meaning).

Given this phenomenon, we should expect that keywords in context searches can often reflect an incomplete picture of a term's ordinary meaning. Because legal corpus linguistics reflects the pragmatics of language *use*, there are a number of uses that are entirely consistent with ordinary meaning that nevertheless *should not* appear frequently in the corpus.[252]

B. Fallacies of Interpretation

This section identifies fallacies in the use of legal corpus linguistics and dictionaries, fallacies made clear by the experimental findings. Of course, there are many other important critiques that are not discussed

---

[251] *But see* Andrew C. Connolly, Jerry A. Fodor, Lila R. Gleitman & Henry Gleitman, *Why Stereotypes Don't Even Make Good Defaults*, 103 COGNITION 1, 13 (2007) (discussing empirical data suggesting that conceptual combinations do not always inherit stereotypes associated with the general concepts). This phenomenon finds some support in the cognitive science of default interpretations. Terms and propositions often have default, presumptive, or preferred interpretations. These default interpretations are often more easily available and are generated in a shorter time than nondefault interpretations. *See generally* STEPHEN C. LEVINSON, PRESUMPTIVE MEANINGS: THE THEORY OF GENERALIZED CONVERSATIONAL IMPLICATURE (2000).

[252] There are, of course, other possible explanations for the pattern of results. I am thankful to Daniel Keller for raising one explanation that is particularly thoughtful. The suggestion is that, when corpus-condition participants evaluated the meaning of "ailac," they were inclined (given the data presented to them) to construct a basic-level category (like car) rather than a superordinate category (like vehicle). One reason to question whether this explains the results is that several entities were categorized as vehicles by over 50% of participants. If the experimental paradigm strongly suggested a basic-level category, we would not expect participants to evaluate many of the entities as vehicles. Nevertheless, this is an intriguing hypothesis and might explain some of the variation between the corpus condition and the other two. This suggestion also raises an interesting question concerning the real-world use of legal corpus linguistics. Perhaps, in some cases, corpus linguistics data are (mistakenly) taken to support that the ordinary meaning of the relevant term is a narrow basic-level category rather than a superordinate category.

here.[253] This section elaborates critiques grounded in the novel experimental results of this Article.

Consider several fallacies in the use of legal corpus linguistics and dictionaries. These are argumentative or inferential errors in common uses of dictionaries and legal corpus linguistics. Individually, these fallacies present significant challenges to common methods of interpretation; collectively, they threaten the plausibility of relying heavily upon only these tools in interpretation.

First consider some fallacies of legal corpus linguistics:

1. The Nonappearance Fallacy: the nonappearance of some use in the corpus indicates that this use is outside of the ordinary meaning.

It is tempting to think that any acceptable use must be found *somewhere* in a large corpus, and any use that is not reflected is therefore not part of the ordinary meaning.[254] Defenders of legal corpus linguistics have suggested this argument with respect to airplanes being vehicles:

> With respect to the use of *vehicle* to reference *airplane*, the answer is simpler. . . . [W]e were unable to find a single collocation or concordance line that reflected the use of *vehicle* to mean *airplane*. . . . [B]ased on its absence from any of our corpus data, we might ask if *airplane* is even a possible sense of *vehicle*.[255]

The experimental results here suggest this argument moves too quickly. Legal corpus linguistics often neglects nonprototypical uses of a term. A concordance search for "vehicle" returns predominantly uses involving cars. But this does not mean that only these more prototypical uses reflect the ordinary meaning of "vehicle." As the experimental results (and common sense) indicate, golf carts, airplanes, and horse-drawn carriages are also within the modern ordinary meaning of "vehicle."

It is important to recognize this fallacy in practice, as the argument often seems to have great rhetorical strength: "In an *entire* corpus, containing tens of thousands of uses, there were *none* reflecting such a meaning. Therefore, that meaning is not the ordinary meaning." This argument is fallacious when interpreting modern texts. It is also fallacious in historical interpretation. A historical corpus is often smaller than modern ones, containing even fewer uses (and thus fewer nonprototypical ones).

A second fallacy follows from the same set of observations and experimental results.

---

[253] *See generally, e.g.*, Solan & Gales, *supra* note 44 (arguing that there are various conditions for the helpful use of linguistic corpora in legal analysis).

[254] *See, e.g.*, Carpenter v. United States, 138 S. Ct. 2206, 2238–39 (2018) (Thomas, J., dissenting) (noting that "[t]he phrase 'expectation(s) of privacy' does not appear in . . . the papers of prominent Founders, early congressional documents and debates, collections of early American English texts, or early American newspapers" (footnotes omitted)).

[255] Lee & Mouritsen, *supra* note 14, at 844.

2. The Uncommon Use Fallacy: the relative rarity of some use in the corpus indicates that this use is outside of the ordinary meaning.

Insofar as legal corpus linguistics data may not adequately reflect nonprototypical uses, one cannot conclude that the rarity of use implies that such a use is not part of the term's ordinary meaning. For example, just because "car" appears more often as a vehicle in the corpus than does "bicycle" or "cement-mixer" does not mean that the latter two *clearly* fall outside of the ordinary meaning.

Next consider the comparative use fallacy.

3. The Comparative Use Fallacy: when considering two possible senses, the comparatively greater support for one sense in the corpus indicates that this sense is a better candidate for ordinary meaning.

This fallacy arises when users of legal corpus linguistics aim to determine which of two possible senses is the better candidate for ordinary meaning. This may happen, for example, if there is debate over whether a term is ambiguous: if one possible sense is much more often reflected in the corpus, one might conclude that that sense reflects the (only) plain meaning or the "best" meaning.

However, this too is a fallacious argument. Recall the experimental findings. Ordinary people, law students, and United States judges classified several entities as vehicles when considering ordinary language, but not when considering the corpus data (for example, airplanes, bicycles, electric wheelchairs). Imagine there was a debate over the meaning of "vehicle." Sense-1 is the more inclusive sense (entities like cars and trucks; and also ones like airplanes, bicycles, or electric wheelchairs), and Sense-2 is the more exclusive sense (only cars and trucks; and not airplanes, bicycles, or electric wheelchairs). Users of legal corpus linguistics might be inclined to argue that Sense-2 is the better candidate, as it has more support from the corpus. However, as discussed previously, the omission of nonprototypical uses from the exclusive sense does not mean it is a better sense or one that reflects *the* (only) plain meaning of "vehicle." Just because it is more common to use "vehicle" to refer to cars than to airplanes does not mean it is clear that the ordinary meaning of "vehicle" in any legal text is Sense-2.

These three fallacies — the Nonappearance Fallacy, the Uncommon Use Fallacy, and the Comparative Use Fallacy — each present an individual challenge to common interpretive arguments grounded in legal corpus linguistics data. But we should also note that these three arguments together threaten much of the current usefulness of legal corpus linguistics. If legal corpus linguistics cannot reliably account for omitted or rare uses (from ordinary meaning) or determine which of two possible

senses is more credible, then the method loses much of its promise as a solution to questions of ordinary meaning.

Now consider two fallacies of dictionary use:

4. The "It Fits the Definition" Fallacy: when considering whether a use falls under the ordinary meaning, we should conclude that the use is part of the ordinary meaning if it fits the relevant definition.

In the studies presented here, dictionary users categorized as vehicles several items that were not judged to be vehicles by ordinary language users. Often, dictionary definitions seem to aim to convey a comprehensive set of meanings. In defining "vehicle," we must provide a definition that includes cars and trucks but also airplanes, submarines, and mopeds. The definition "a means of carrying or transporting something," is helpful in achieving this. But such a broad definition also applies to many entities that are not understood as vehicles. For example, participants using their ordinary judgments predominantly reported that roller skates, baby-shoulder carriers, and zip lines are *not* vehicles.[256] Yet many dictionary users categorized these as vehicles.[257] Thus, while it might seem that a dictionary definition is tied tightly to ordinary meaning, the connection is much looser. The mere fact that a use "fits" a dictionary definition (or several definitions) does not imply that the use is consistent with ordinary understanding. Given the practical nature of dictionaries, which aim to define succinctly a broad range of meaningful uses, we should expect that some definitions might appear to apply more broadly than the ordinary meaning.

At the same time, and perhaps surprisingly, there may be particular features of a dictionary definition that seem to exclude certain uses from ordinary meaning.

5. The "It Doesn't Fit the Definition" Fallacy: when considering whether a use falls under the ordinary meaning, we should conclude that the use is not part of the meaning if it does not fit the relevant definition.

Sometimes dictionaries include features that are common features of category members but not necessary criteria of category membership. This is especially common in multipart dictionary definitions. For example, perhaps "cruel" punishment is often, *but not necessarily*, characterized by the infliction of pain for pain's sake. Or perhaps a vehicle is typically, *but not necessarily*, mobile.[258] It is sometimes a mistake to

---

[256] *See supra* Figure 5.

[257] *See supra* Figure 6.

[258] *See supra* Figures 5 & 6 and accompanying text (commenting that ordinary concept participants were divided roughly evenly over whether a nonmobile World War II truck was a vehicle, but that dictionary participants overwhelmingly disagreed).

point to a particular aspect of one dictionary definition and argue that any use that does not meet that criterion cannot be part of the ordinary meaning.

C. An Empirical Challenge to Formalist, Textualist, and Originalist Interpretation

This section considers the experimental implications for a certain set of interpretive theories, namely those that *rely* on dictionary definitions or legal corpus linguistics data to determine the "ordinary meaning" of terms and phrases and resolve legal disputes via that ordinary meaning. This section argues that, in light of the data, the argumentative burden shifts to these theories to provide a reliable and nonarbitrary methodology.

Some theories of interpretation assume an "ordinary meaning" of terms and phrases, take that meaning as the criterion of legal interpretation, and rely on empirical evidence — for example, a dictionary definition — to determine the meaning and legal outcome. Consider Professors Alan Schwartz and Robert E. Scott's characterization of common formalist and contextualist approaches to contract interpretation:

> Contests over the meaning of contract terms thus follow a predictable pattern: one party claims that the words in a disputed term should be given their standard dictionary meaning, as read in light of the contract as a whole, the pleadings, and so forth. The counterparty argues either that the contract term in question is ambiguous and extrinsic evidence will resolve the ambiguity, or that extrinsic evidence will show that the parties intended the words to be given a specialized or idiosyncratic meaning that varies from the meaning in the standard language.[259]

Here the first party's approach reflects a common formalist approach to contract interpretation. If the plain meaning of the contract is unambiguous, the ordinary meaning of a term — perhaps evinced only by its dictionary definition — suffices in determining the interpretive outcome.

A similar approach is common in textualist and originalist statutory and constitutional interpretation. Professor Victoria Nourse has documented the increasing tendency of textualist interpreters to rely on the ordinary meaning of specific words in statutes through textual "gerrymandering,"[260] "intense decontextualization,"[261] and "reducing the statute's meaning to a particular word or two."[262] Moreover, as Professors

---

[259] Alan Schwartz & Robert E. Scott, *Contract Interpretation Redux*, 119 YALE L.J. 926, 962 (2010).

[260] Nourse, *supra* note 32, at 681.

[261] *Id.* at 669.

[262] *Id.* at 681.

Abbe Gluck and Lisa Bressman note, these interpretive tasks often concern very ordinary words.[263] And, in many cases, both "liberal" and "conservative" Justices rely on dictionary definitions to establish the ordinary meaning of these important terms.[264]

Insofar as the use of dictionary definitions and legal corpus linguistics is meant to identify how ordinary people would understand particular terms and phrases, the experimental results here indicate that both tools are surprisingly unreliable in that task. As section IV.E estimated, each method diverges from ordinary understanding in the range of at least 20–35%, and in some cases over 80%.

More broadly, on these theories, dictionary definitions and legal corpus linguistics should track the same fact: *the* "ordinary meaning." The extreme divergences between the use of dictionary definitions and legal corpus linguistics — by ordinary people, law students, and judges — provides further cause for concern. If interpretive theories posit an "ordinary meaning" that serves as a primary criterion of legal interpretation, we must know much more from those theories about how, precisely, use of dictionaries and legal corpus linguistics can reliably achieve that task.[265]

The studies also provide evidence concerning the search for historical ordinary meaning, or "original public meaning." Insofar as there is no compelling reason to think that these tools perform better in historical analysis, the results provide evidence that they are unreliable in historical interpretation:

1. Empirical Claim: The modern uses of dictionaries and legal corpus linguistics do not accurately reflect people's ordinary judgments.
2. Reliability Premise: A method that does not accurately reflect people's judgments is not a reliable method of determining ordinary meaning.
3. Intermediate Conclusion: The empirical results provide evidence that the method is unreliable in modern interpretation.

---

[263] Gluck & Bressman, *Part I*, *supra* note 45, at 955.

[264] *See* Krishnakumar, *supra* note 24, at 250–51.

[265] For example, in a thoughtful response, Neal Goldfarb takes issue with my conclusion that the experimental results show legal corpus linguistics to be inaccurate in determining ordinary meaning. He argues that the experiments assume a particular conception of what constitutes ordinary meaning and that caselaw reflects the existence of other conceptions. Under one of those conceptions, argues Goldfarb, ordinary meaning *consists* in the way that the word or other expression is ordinarily used. Goldfarb also points to a conception of ordinary meaning as consisting in how the word or expression is ordinarily understood, arguing that although my methodology also appeals to the idea of ordinary understanding, there are multiple conceptions of that idea, just as there are multiple conceptions of ordinary meaning. He asserts that while my methodology is appropriate under one of the conceptions of ordinary meaning, it is not appropriate under the other. Neal Goldfarb, *Varieties of Ordinary Meaning: Comments on Kevin P. Tobia,* Testing Ordinary Meaning, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3553016 [https://perma.cc/32FZ-TGNH].

4. Historical Inference: In the absence of historically distinguishing factors, evidence of a method's unreliability in modern interpretation also serves as evidence about that method's unreliability in historical interpretation.
5. Conclusion: The results provide evidence that the methods are unreliable in historical interpretation.

Like the conclusion for modern interpretation, this conclusion shifts the argumentative burden to theories that rely upon these tools to elaborate and justify their methodologies. The two keys features of this challenge are nonarbitrariness and demonstrable reliability.

First, consider nonarbitrariness. There are many choices one must make in interpretation. For example, for interpretive theories that advocate using dictionaries, *which* dictionaries should be used? Relying on different dictionaries for different cases invites a charge of arbitrariness. The sources of arbitrariness are even broader for legal corpus linguistics: Exactly how many searches will be conducted, what precisely will be searched and how will the search string be determined, and what number or percentage of conforming uses will "count" as an instance of ordinary or public meaning?

Although this paper has largely set these questions aside, a defense of an interpretive methodology relying on dictionaries or legal corpus linguistics must address these fundamental concerns. But there are also new sources of arbitrariness illuminated by the experimental results. Principally, consider the arbitrariness in choosing to use dictionaries *or* legal corpus linguistics.

The experimental results suggest that legal corpus linguistics and dictionaries are not just sometimes divergent; they often provide strongly opposing verdicts about ordinary meaning.[266] Insofar as a theorist or jurist endorses dictionaries in one instance and legal corpus linguistics in another — with no further supporting reasons — this raises a new question of arbitrariness. The same judges or interpreters sometimes point out the absence of (nonprototypical) uses in a corpus in one case but rely on the breadth of a dictionary definition in another.[267] Resolving this apparent arbitrariness becomes more pressing where the choice of methodology seems to match the desired political or legal outcome.

Thus, there are burdens on theorists who rely upon dictionaries and corpora to elaborate and defend a nonarbitrary use of their tools. Reporting dictionary definitions and detailed corpus data often conveys

---

[266] *See supra* section IV.E.7, pp. 772.

[267] *Compare, e.g.*, Smith v. United States, 508 U.S. 223, 241–44 (1993) (Scalia, J., dissenting) (arguing that a broad dictionary definition of "use" should be rejected and instead relying on prototypical examples of "use" in ordinary language to argue that the ordinary meaning is narrower than the dictionary definition), *with* District of Columbia v. Heller, 554 U.S. 570, 581–82 (2008) (opinion of Scalia, J.) (arguing that "arms" should be construed exclusively on the basis of dictionary definitions).

an impression of legitimacy and scientific rigor. However, these values are illusory if the method of interpretation is subtly (consciously or unconsciously) altered in each case.

The first burden can be satisfied with sufficiently detailed theory and methods. Interpreters must simply commit to a list of interpretive choices. For example, perhaps the first definition of a term in *X* dictionary is deemed the authoritative source in contract interpretation.

The second burden, to articulate a demonstrably reliable use of these tools, requires empirical study. If dictionary definitions and legal corpus linguistics methods are unreliable, it does not matter much that they are applied systematically. We can construct many nonarbitrary methods of interpretation, but any such method is unconvincing until it is also shown (empirically) to be reliable.

The burden now rests with theories that rely on these tools of discovering ordinary meaning. We should remain open to the very real possibility that such a challenge might be met. But, for a moment, imagine that such a theory of interpretation does not adequately meet this burden. How should the theory fare?

Recall the divergence or "error" rates for dictionaries and legal corpus linguistics.[268] These rates were computed by taking on a core assumption of textualist and originalist theories that use dictionaries and legal corpus linguistics: there is an empirical *fact* about ordinary meaning, grounded in what language communicates to ordinary people. Assuming, for the sake of argument, that it is sensible to think that there is some such "ordinary meaning," we computed how often dictionaries and legal corpus linguistics accurately reflect that fact.

Overall, the error rate for one relying on each method was between 20–35%. In many cases, the rate was larger: 50%, 75%, even 100%. These numbers may seem abstract, but consider what they represent: the data suggest that judges relying on legal corpus linguistics and dictionary definitions would arrive at the wrong interpretation (by their own theory's lights) once in every three to five cases, and perhaps even more frequently.

D. Insights for Interpretive Theories Uncommitted to "Ordinary Meaning"

Seventy years ago, Justice Frankfurter described the difficulty of legal interpretation: "Anything that is written may present a problem of meaning . . . . The problem derives from the very nature of words."[269] To be sure, contracting parties or legislative drafters can reduce some potential uncertainty with careful drafting, but inevitably "[a] large area is bound to remain."[270]

---

[268] *See supra* section IV.E.9, pp. 773–76.
[269] Frankfurter, *supra* note 13, at 528.
[270] *Id.*

For theories and theorists uncommitted to, or even skeptical of, the notion of a single "ordinary meaning," the experimental results here might be taken to support this Frankfurtian perspective. Despite the promise of well-researched dictionaries and large, data-driven legal corpus linguistics analyses, some hard problems of meaning inevitably remain.

The results illuminate two different criteria that are often relevant in assessments of legal texts: a more extensive criterion and a more narrow, prototypical criterion. For example, when we consider the meaning of "vehicle" in a statute or insurance contract,[271] an extensive criterion indicates that airplanes, canoes, and even drones are vehicles, while a prototypical criterion indicates that these entities are not vehicles. The findings show that dictionaries or legal corpus linguistics sometimes reflect *one* of these criteria — often dictionaries reflect the extensive criterion and legal corpus linguistics the prototypical one — but the question remains: Which of these — if either — *should* serve as a criterion in legal interpretation?

We might consider these two criteria against common values and aims of legal interpretation. Most of those values do not count strongly in favor of one criterion over the other. Is it more likely that using a prototypical criterion in contract interpretation will capture the parties' intent? Not necessarily — as the experiments indicate, in some cases people will understand a term to apply more broadly than to only more prototypical uses. Is it more likely that using an extensive criterion in statutory interpretation will lead to more robustly "public" laws?[272] Not necessarily — as the experiments indicate, in some cases people will not understand a term to apply as broadly as an extensive criterion suggested by a dictionary.

Ultimately, it is unlikely that either of these criteria should serve universally as the criterion of interpretation. As Justice Scalia put it: "A text should not be construed strictly, and it should not be construed leniently; it should be construed reasonably, to contain all that it fairly means."[273]

The previous section discussed certain formalist, textualist, and originalist theories that often operate as if dictionaries and legal corpus linguistics deliver such a meaning — not strict, not lenient, but simply *ordinary*. It concluded that — in light of the experimental findings revealing dramatic divergences among use of dictionaries, use of legal corpus linguistics, and ordinary understanding — those views have the

---

[271] *See supra* section I.B, pp. 739–42.

[272] *See, e.g.*, Antonin Scalia, *The Rule of Law as a Law of Rules*, 56 U. CHI. L. REV. 1175, 1178–80 (1989) (arguing that judges should abide by clear principles in statutory interpretation in order to supply predictability for the public).

[273] Scalia, *supra* note 59, at 23.

burden to articulate and demonstrate *how* such a task should be achieved.

But on a range of other plausible interpretive theories, the experimental findings about dictionaries and legal corpus linguistics provide constructive insight. On many interpretive views, there are certain circumstances in which a text *should* be construed strictly or leniently. Here I consider three such circumstances: ones triggered by (a) applicable canons of interpretation, (b) relevant context, or (c) the text's purpose.

First, consider interpretive canons. In both contractual and statutory interpretation, the *ejusdem generis* canon holds that "the meaning of a word in a series of words is determined 'by the company it keeps'"[274]: when a general word or phrase follows a list of specifics, that general word or phrase should be interpreted to include just those of the same type listed. Legal corpus linguistics data about prototypical uses could serve as useful evidence confirming inclusion under *ejusdem generis*. For example, finding that "vehicle" refers to buses would be evidence in favor of interpreting a statute concerning "cars, trucks, and other vehicles" to include buses.

In some other circumstances, use of both dictionaries and legal corpus linguistics would be instructive. For example, consider criminal contexts in which the rule of lenity applies.[275] Insofar as dictionaries and corpora provide evidence about different senses of a term in this context (for the sake of simplicity, say a "prototypical sense" and an "extensivist sense"), one might want to compare both senses and apply whichever is more consistent with the rule of lenity. Depending on the context, either the more extensive or more prototypical sense could comport with the rule of lenity.

Second, many plausible theories of interpretation look to the context — the full contractual text in which the disputed contract term or clause is embedded, or the whole act within which the relevant statutory term or clause is embedded. This includes some sophisticated forms of textualism and originalism — on those views, a dictionary definition or legal corpus linguistics dataset concerning a single term or phrase would often not be sufficient to determine the interpretation.

For example, consider again the "no vehicles in the park" example. Although that rule does not provide much context that implies the appropriateness of an extensivist or prototypical sense of "vehicle," a modified version might provide that information. "*Any and all* vehicles are prohibited from the park" might suggest that "vehicles" should be construed rather extensively. Alternatively, "*only* cars, trucks, and other

---

[274] 242-44 E. 77th St., LLC v. Greater N.Y. Mut. Ins. Co., 815 N.Y.S.2d 507, 510 (App. Div. 2006) (quoting People v. Illardo, 399 N.E.2d 59, 63 (N.Y. 1979)); *see also, e.g.*, Gooch v. United States, 297 U.S. 124, 128 (1936); Aspen Advisors LLC v. United Artists Theatre Co., 861 A.2d 1251, 1265 (Del. 2004).

[275] For example, there is a textual ambiguity and neither of the two possible senses is inconsistent with legislative intent.

vehicles are prohibited from the park" might suggest a more prototypical sense of "vehicles."

In practice, using both dictionaries and corpora is likely better than relying on either alone. For example, imagine that the interpretive context calls for a broad, extensivist reading of the term or phrase. While dictionaries are a comparatively better source for generating this extensivist sense, we might also cross-check the corpus for relatively rare uses. Although the absence of a use from a corpus cannot guarantee that such a use is outside of the ordinary meaning, the presence of a use from a corpus can provide some evidence that a use is within at least a nonprototypical sense of the ordinary meaning.[276]

As a final example, consider the significance of a text's purpose. A theory that takes purpose as a relevant interpretive criterion might look to either the prototypical or broad sense of meaning. Keeping with the example of "vehicles," if the purpose of a statute is to register widely any means of transportation, the broad criterion indicated by a dictionary would be more instructive.[277] Conversely, if a contract's purpose is to provide limited insurance for the use of common "vehicles," the prototypical criterion indicated by the legal corpus linguistics data may be more instructive.[278]

These considerations about canons, context, and purpose indicate that dictionary definitions and legal corpus linguistics data can be useful inputs into legal interpretive analyses. Yet these measures may not be equipped *in principle* to deliver simple and unequivocal answers to inquiries about the ordinary meaning of legal texts. Legal interpreters will have to look beyond the simple dictionary definition and corpus frequency analysis — to the legal text's context, history, and purpose; and to their other interpretive commitments.

---

[276] Even here, caution is required. For example, it is possible to find metaphorical or sarcastic uses in the corpus, which would not provide such evidence.

[277] ESKRIDGE, *supra* note 17, at 9 ("Text and purpose are like the two blades of a scissors; neither does the job without the operation of the other."); ESKRIDGE ET AL., *supra* note 59, at 6 (noting that although "prototypical" meaning has an important role to play in statutory interpretation, judges frequently adopt a more "extensive" meaning, to give effect to the statute's purpose).

[278] W.W.W. Assocs., Inc. v. Giancontieri, 566 N.E.2d 639, 642 (N.Y. 1990) (stating that the court reads the agreement "as a whole to determine its purpose and intent"); RESTATEMENT (SECOND) OF CONTRACTS § 202(1) (AM. L. INST. 1981) ("Words and other conduct are interpreted in the light of all the circumstances, and if the principal purpose of the parties is ascertainable it is given great weight.").

CONCLUSION

This Article has developed a method of testing two of the fundamental tools of ordinary meaning analysis — dictionary definitions and patterns of word usage through legal corpus linguistics.[279] A series of experiments examined judgments of ordinary people, law students, and United States judges, providing evidence bearing on the process and reliability of dictionary and corpus use in interpretation.

The results help identify several common fallacies of interpretation. As one example, recall the "Nonappearance Fallacy," the mistaken assumption that the nonappearance of some use in a corpus indicates that this use is outside of ordinary meaning. As the experimental results indicate, ordinary meaning sometimes diverges from ordinary use: people's understanding of language is not always reflected in recorded speech and writing, especially their understanding concerning nonprototypical category membership.

For certain textualist and originalist views that are committed to the existence of a single ordinary meaning of terms like "vehicle" and phrases like "carrying a firearm," the data suggest that popular methods of dictionary use and legal corpus linguistics carry serious risks of diverging from ordinary understanding. These results shift the argumentative burden to theorists and practitioners who rely on these tools to determine legal outcomes: in light of the data, these views must articulate and demonstrate a nonarbitrary and reliable method of interpretation.

Finally, from the perspective of interpretive theories that are uncommitted to, or even skeptical of, the notion of a single "ordinary meaning" that determines legal outcomes across a range of cases and contexts, the findings illuminate two different criteria that are often relevant in assessing the meaning of legal texts: a more extensive criterion and a narrower, more prototypical criterion. Although dictionaries and legal

---

[279] For other recent examples of such empirically informed and experimental jurisprudence, see Kevin P. Tobia, *Experimental Jurisprudence*, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3680107 [https://perma.cc/7C7V-8MZN]. *See also* Ivar R. Hannikainen & Raff Donelson, *Fuller and the Folk: The Inner Morality of Law Revisited*, *in* 3 OXFORD STUDIES IN EXPERIMENTAL PHILOSOPHY 6 (Tania Lombrozo et al. eds., 2020) (on the ordinary concept of law); Markus Kneer & Sacha Bourgeois-Gironde, *Mens Rea Ascription, Expertise and Outcome Effects: Professional Judges Surveyed*, 169 COGNITION 139 (2017) (on the ordinary concept of intentional action); Macleod, *supra* note 54 (on the ordinary concept of causation); Christian Mott, *Statutes of Limitations and Personal Identity*, *in* 2 OXFORD STUDIES IN EXPERIMENTAL PHILOSOPHY 243 (Tania Lombrozo et al. eds., 2018) (on the ordinary concept of identity); Kevin P. Tobia, *How People Judge What Is Reasonable*, 70 ALA. L. REV. 293 (2018) (on the ordinary concept of reasonableness); Roseanna Sommers, *Commonsense Consent*, 129 YALE L.J. 2232 (2020) (on the ordinary concept of consent); Joshua Knobe & Scott Shapiro, *What Cognitive Science Can Teach Us About Proximate Causation*, U. CHI. L. REV. (forthcoming) (on file with the Harvard Law School Library) (on the ordinary concept of causation); Kevin P. Tobia, Essays in Experimental Jurisprudence (2019) (unpublished Ph.D. dissertation, Yale University) (on file with the Harvard Law School Library).

corpus linguistics can help us assess these criteria, a hard legal-philosophical question remains: Which of these two criteria *should* guide the interpretation of terms and phrases in legal texts? Insofar as there is no compelling case to prefer one, the results suggest that dictionary definitions, legal corpus linguistics, or even other more scientific measures of meaning may not be equipped *in principle* to deliver simple and unequivocal answers to inquiries about the ordinary meaning of legal texts.