

Testing Ordinary Meaning: An Experimental Assessment of What Dictionary Definitions and Linguistic Usage Data Tell Legal Interpreters

Kevin P. Tobia

133 HARV. L. REV. (forthcoming 2020)

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.” Legal interpretation—of contracts, statutes, wills, trusts, deeds, patents, regulations, treaties, and constitutions—regularly includes evaluation of how ordinary people would understand the text. Theorists and practitioners often treat the search for ordinary meaning as an empirical inquiry, aiming to discover facts about how ordinary people would understand language. To discover ordinary meaning, interpreters increasingly recommend as evidence a relevant term’s *dictionary definition* or its *pattern of usage* across various sources in an English-language corpus. However, the most central question about these sources of evidence remains open: Do these popular methods accurately reflect ordinary meaning?

To assess this question, this paper develops and employs a novel method of “experimental jurisprudence.” A series of experimental studies (N = 4,162) reveals systematic divergences among the verdicts delivered by modern concept use, dictionary use, and corpus linguistics use. For example, today people apply the concept of a vehicle differently from the way in which they apply modern dictionary definitions or modern corpus linguistics data concerning vehicles. The same results arise across levels of legal expertise—participants included 230 “elite-university” law students (e.g. at Harvard and Yale) and 98 United States judges—and across various terms and phrases, including “vehicle,” “labor,” “weapon,” “carrying a firearm,” and “tangible object.”

The paper elaborates several implications of these results. First, the results provide insight into what dictionaries and corpus linguistics suggest to legal interpreters. Drawing on insights from linguistics and psychology, I distinguish between “prototypical” and “broad” senses of the same term. For example, a car is a prototypical vehicle, while airplanes, bicycles, and canoes are less prototypical vehicles. An extensive criterion would include all of those entities as vehicles, while a prototypical criterion would include only cars. This distinction about language is well-known, but the experiments show that the distinction also illuminates ordinary meaning’s sources of evidence. That is, dictionaries and corpus linguistics often track only *one* of these criteria—dictionaries tend to track the broad criterion and corpus linguistics the prototypical one.

Second, I identify several fallacies of interpretation that are supported by the results. As one example, consider “The Non-Appearance Fallacy,” the mistaken assumption that the non-appearance 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 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, ordinary meaning sometimes diverges from ordinary use: People’s full understanding of language is not always reflected in recorded speech and writing, especially their understanding concerning non-prototypical category membership.

Third, I evaluate the findings' significance for different theories of legal interpretation. First I consider certain formalist, 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," one which is outcome-determinative without reference to further context, textual purpose, or even type of law (e.g. criminal vs. contract). The data suggest that popular methods of dictionary-use and corpus linguistics carry serious risks of diverging from ordinary understanding—conservatively estimated, 20-35%. And in some circumstances, even judges' use of these methods carried extremely large divergence rates—between 80-100%. The results shift the argumentative burden to theorists and practitioners that rely on these tools to determine legal outcomes: In light of the data, these views must articulate and demonstrate a reliable method of interpretation.

Finally, I consider the 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 findings illuminate two different criteria that are often relevant in interpretation: a more extensive criterion and a more narrow, prototypical criterion. Although dictionaries and 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, 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.

CONTENTS

INTRODUCTION	4
I. ORDINARY MEANING	11
A. <i>The Concept of Ordinary Meaning</i>	11
B. <i>Theories of Interpretation</i>	13
II. SOURCES OF INTERPRETIVE EVIDENCE	15
A. <i>Individual Intuition</i>	15
B. <i>Dictionary Definitions</i>	17
C. <i>Linguistic Usage Data via Corpus Linguistics</i>	17
III. ORDINARY MEANING THEORY’S EMPIRICAL ASSUMPTIONS	18
A. <i>Empirical Critiques</i>	19
B. <i>Are Dictionaries and Corpus Linguistics Reliable? A Thought Experiment</i>	20
IV. AN EXPERIMENTAL TEST OF DICTIONARIES AND CORPUS LINGUISTICS	22
A. <i>Experiment 1</i>	24
B. <i>Experiment 2: The Process of Using Dictionaries and Corpora</i>	27
C. <i>Experiment 3: Expert Judges</i>	29
D. <i>Experiment 4: Replication Across Ten Examples</i>	31
E. <i>Summary and Interpretation</i>	33
1. <i>Judges and Non-Experts Judge Meaning Similarly</i>	33
2. <i>Ordinary Meaning is Often Unclear</i>	34
3. <i>Judges and Non-Experts Apply Dictionaries Similarly</i>	34
4. <i>Judges and Non-Experts Apply Corpus Linguistics Similarly</i>	35
5. <i>Corpus Linguistics Use Reflects Narrow, Prototypical Uses</i>	35
6. <i>Dictionary-Use Can Reflect Extensive Uses</i>	36
7. <i>Dictionaries and Corpus Linguistics Provide Divergent Verdicts</i>	37
8. <i>Dictionary and Corpus Linguistics Verdicts Diverge from Ordinary Judgment</i> ..	38
9. <i>On Average, Corpus Linguistics and Dictionaries Had 20-35% Error Rates</i>	38
10. <i>In Some Circumstances, Error Rates Reached 80-100%</i>	40
V. HOW JURISTS USE DICTIONARIES AND CORPUS LINGUISTICS	42
A. <i>Dictionaries Often Supply “Broad” Senses and Corpus Linguistics “Narrow” Ones</i> ..	42
B. <i>Political Uses of Dictionaries and Corpus Linguistics</i>	47
VI. IMPLICATIONS	50
A. <i>Understanding the Use of Dictionaries and Corpora in Interpretation</i>	51
B. <i>Fallacies of Interpretation</i>	54
C. <i>An Empirical Challenge to Formalist, Textualist, and Originalist Interpretation</i>	57
D. <i>Insights for Interpretive Theories Uncommitted to “Ordinary Meaning”</i>	59
CONCLUSION.....	61

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. Across the interpretation of contracts,¹ wills,² trusts,³ deeds,⁴ patents,⁵ statutes,⁶ regulations,⁷ treaties,⁸ and constitutions,⁹ 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: “every legal system recognizes the importance of ordinary meaning.”¹⁰ This is for good reason: “What method of ... interpretation would view the ordinary meaning of words as completely irrelevant?”¹¹ Even legal theorists who advocate looking beyond ordinary meaning acknowledge that, in interpretation “one certainly begins there.”¹²

Despite this general agreement concerning ordinary meaning's legal relevance, there is significant debate about how exactly to elaborate the concept of ordinary meaning. Broadly speaking, 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 when 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. Various legal debates concern this original or historical ordinary meaning of a text, especially in statutory and constitutional contexts,¹³ but also in contractual ones.¹⁴

¹ E.g. CAL. CIV. CODE § 1644 (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 (Fed. Cir. 2000) (“We give the words of the agreement their ordinary meaning unless the parties mutually intended and agreed to an alternative meaning”).

² E.g. CAL. PROB. CODE § 21122 (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”).

³ E.g. *id.*

⁴ E.g. *Lambert v. Pritchett*, 284 S.W.2d 90, 91 (Ky. 1955) (“terms are to be construed and understood according to their plain, ordinary, and popular sense”); *Burdette v. Bruen*, 191 S.E. 360 (W. Va. 1937).

⁵ E.g. *Phillips v. AWH Corp.*, 415 F.3d 1303 (Fed. Cir. 2005) (“the ordinary meaning of a claim term is its meaning to the ordinary artisan”).

⁶ E.g. *Moskal v. United States*, 49 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.”).

⁷ E.g. *Bowles v. Seminole Rock & Sand Co.*, 325 U.S. 410, 414 (1945) (the court's tools “are the plain words of the regulation and any relevant interpretations of the Administrator”); *see also Christensen v. Harris County* 529 U.S. 576 (2000).

⁸ E.g. Convention on the Law of Treaties art. 31, May 23, 1969, 1155 U.N.T.S. 331 [hereinafter Vienna Convention] (“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 also Curtis J. Mahoney, Treaties as Contracts: Textualism, Contract Theory and the Interpretation of Treaties*, 116 YALE L.J. 824 (2017).

⁹ *See, e.g., Lawrence B. Solum, The Constraint Principle: Original Meaning and Constitutional Practice* (Apr. 11, 2018).

¹⁰ Daniel A. Farber, *Hermeneutic Tourist: Statutory Interpretation in Comparative Perspective*, 81 CORN. L. REV. 513, 516 (1986) (reviewing INTERPRETING STATUTES: A COMPARATIVE STUDY (D. Neil MacCormick & Robert S. Summers, 1991) (citing the relevance of ordinary meaning to legal interpretation in jurisdictions including Argentina, Germany, Finland, France, Italy, Poland, Sweden, the U.K., and the U.S.)).

¹¹ *Id.* at 516 (citing Edward Rubin, Book Review, 41 AM. J. COMP. L. 128, 139).

¹² Felix Frankfurter, *Some Reflections on the Reading of Statutes*, 47 COLUM. L. REV. 527, 528 (1947).

¹³ Debate is especially divisive in the statutory and constitutional context. *See e.g., Victoria Nourse, Reclaiming the Constitutional Text from Originalism: The Case of Executive Power*, 106 CALIF. L. REV. 1 (2018); Lawrence B. Solum, *The Constraint Principle: Original Meaning and Constitutional Practice* (Apr. 11, 2018). Following Thomas R. Lee & Stephen C. Mouritsen, *Judging Ordinary Meaning*, 127 YALE L.J. 788, 825-826 (2018), I use “original public meaning” to refer to a legal text's communicative content (or “ordinary meaning”) at the relevant time. For example, the original public meaning of a 1967 statute is that text's ordinary meaning in 1967.

¹⁴ E.g. Stephen C. Mouritsen, *Contract Interpretation with Corpus Linguistics*, 94 WASH. L. REV. (2019).

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,¹⁵ where ordinary meaning is “the most fundamental principle.”¹⁶ Analysis of the ordinary, plain, or natural meaning underlies popular approaches to contract interpretation,¹⁷ in the U.S. and abroad,¹⁸ which remains the “most important” source of contracts litigation.¹⁹ Ordinary meaning analysis also informs interpretation of patents,²⁰ trusts, and wills.²¹

An empirical study of the 2006-2009 Supreme Court term 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.”²² In light of the two most recent Supreme Court appointments, ordinary meaning analysis will likely continue to hold a significant place.²³ 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. Dictionaries may not provide authoritative, binding interpretation 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.²⁴

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.²⁵

This emphasis on ordinary meaning is typically associated with conservative legal thought. But the consideration of ordinary meaning is in fact broader. In *Muscarello v. United States* Justice Breyer analyzed the statute’s “ordinary” meaning, employing similar methods to those recommended by Justice Gorsuch. Breyer referred to (i)

¹⁵ See, e.g., AHARON BARAK, *PURPOSIVE INTERPRETATION IN LAW* (2005); ROBERT A. KATZMANN, WILLIAM N. ESKRIDGE, JR., *INTERPRETING LAW: A PRIMER ON HOW TO READ STATUTES AND THE CONSTITUTION; JUDGING STATUTE*; ANTONIN SCALIA & BRYAN A. GARNER, *READING LAW: THE INTERPRETATION OF LEGAL TEXTS* (2012); LAWRENCE SOLAN, *THE LANGUAGE OF JUDGES* (1993); LAWRENCE SOLAN, *THE LANGUAGE OF STATUTES: LAWS AND THEIR INTERPRETATION* (2010); BRIAN G. SLOCUM, *ORDINARY MEANING: A THEORY OF THE MOST FUNDAMENTAL PRINCIPLE OF LEGAL INTERPRETATION* (2015); Richard H. Fallon, Jr. *The Meaning of Legal “Meaning” and Its Implications for Theories of Legal Interpretation*, 82 U. CHI. L. REV. 1235 (2015).

¹⁶ SLOCUM, *supra*.

¹⁷ See generally Gregory Klass, *Contract Exposition and Formalism* (2017); 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).

¹⁸ See e.g., *Arnold v. Britton* [2015] UKSC 36 (“the 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”).

¹⁹ Ronald J. Gilson, Charles F. Sabel & Robert E. Scott, *Text and Context: Contract Interpretation as Contract Design*, 100 CORNELL L. REV. 23, 25 (2014).

²⁰ See generally Tun-Jen Chiang & Lawrence B. Solum, *The Interpretation-Construction Distinction in Patent Law*, 123 YALE L.J. 530 (2013).

²¹ E.g. CAL. PROB. CODE § 21122 (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”); NAOMI R. CAHN, ALYSSA DIRUSSO & SUSAN N. GARY, WILLS, TRUSTS, AND ESTATES IN FOCUS 144 (2019).

²² Anita S. Krishnakumar, *Statutory Interpretation in the Roberts Court’s First Era: An Empirical and Doctrinal Analysis*, 62 HASTINGS L.J. 221, 251 (2010).

²³ E.g., Brett M. Kavanaugh, *Review, Fixing Statutory Interpretation*, 120 HARV. L. REV. 2118 (2016).

²⁴ *Id.* at 2150.

²⁵ *Henson v. Santander Consumer USA Inc.*, 137 S. Ct. 1718 (2017).

dictionary definitions and (ii) patterns of word usage in “computerized newspaper databases.”²⁶

Similarly, although ordinary meaning analysis is often associated with textualism and formalism,²⁷ a diverse range of theories endorse the relevance of ordinary meaning in legal interpretation.²⁸ Some of these views represent interpretive pluralism, considering ordinary meaning alongside other criteria in legal interpretation. But even many explicitly non-textualist and non-formalist theories rely on ordinary meaning in an indirect way. When interpreting contracts, aiming to uncover and preserve the party’s *intentions*, “the words of an integrated agreement remain the most important evidence of intention.”²⁹ And in determining the *purpose* of a statute purposivists often ask, “what would a reasonable human being intend this specific language to accomplish?”³⁰ On these views, ordinary meaning is not itself a criterion of legal interpretation, but it is nevertheless important evidence of other interpretive criteria, such as mutual intent or the text’s purpose.

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?³¹

That critique highlights a crucial insight. Ordinary meaning inquiries are often understood as *empirical* ones, which aim to discover descriptive *facts* about meaning.³² 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 “corpus linguistics” to analyze patterns of language usage across a corpus.³³ The popularity of these methods is not difficult to explain. Dictionary-use and corpus linguistics are relatively easy to use. Moreover, they often seem “objective” and even “scientific.”³⁴

Both methods are also *increasingly* popular. Dictionaries are cited more today than ever before within the Supreme Court and Circuit Courts.³⁵ Corpus linguistics is certainly

²⁶ *Muscarello v. United States*, 524 U.S. 125, 128-129 (1998).

²⁷ See Victoria Nourse, *Textualism 3.0: Statutory Interpretation After Justice Scalia*, 70 ALA. L. REV. 667 (2019); Gregory Klass, *Contract Exposition and Formalism* (discussing interpretive formalism in contract law).

²⁸ See, e.g., WILLIAM N. ESKRIDGE, JR. INTERPRETING LAW: A PRIMER ON HOW TO READ STATUTES AND THE CONSTITUTION 35 (2016) (“[t]here are excellent reasons for the primacy of the ordinary meaning rule”); Lee & Mouritsen, *supra* note 13; Lawrence B. Solum, *Surprising Originalism*, CONLAWNOW (2018); see also Elena Kagan, *The Scalia Lecture: A Dialogue with Justice Kagan on the Reading of Statutes*, HARV. L. TODAY (Nov. 17, 2015), <http://today.law.harvard.edu/in-scalia-lecture-kagan-discusses-statutory-interpretation/> (“[W]e’re all textualists now”).

²⁹ RESTATEMENT (SECOND) OF CONTRACTS § 212(b).

³⁰ Stephen Breyer, *On the Uses of Legislative History in Interpreting Statutes*, 65 S. CAL. L. REV. 845, 854; see also HENRY M. HART, JR. & ALBERT M. SACKS, THE LEGAL PROCESS: BASIC PROBLEMS IN THE MAKING AND APPLICATION OF LAW 1182, 1184 (William N. Eskridge, Jr. & Phillip P. Frickey eds., 1994).

³¹ E.g. Richard H. Fallon, Jr. *The Meaning of Legal “Meaning” and Its Implications for Theories of Legal Interpretation*, 82 U. CHI. L. REV. 1235 (2015); Cass Sunstein, *There is Nothing that Interpretation Just Is*, 30 CONST. COMMENT. 193, 194-195 (2015).

³² E.g. Randy E. Barnett, *Interpretation and Construction*, 34 HARV. J.L. & PUB. POL’Y 65 (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 5-14 (1999))).

³³ See Lawrence B. Solum, *Originalist Methodology*, 84 U. CHI. L. REV. 269 (2017).

³⁴ See, e.g., Lawrence Solan, *Can Corpus Linguistics Help Make Originalism Scientific?* 126 YALE L.J. F. 57 (2016).

³⁵ John Calhoun, Note, *Measuring the Fortress: Explaining Trends in Supreme Court and Circuit Court Dictionary Use*, 124 YALE L.J. 484 (2014).

less prevalent, but it also grows in use and esteem.³⁶ The Supreme Court has examined patterns of word use through newspaper databases³⁷ and state supreme courts have searched corpora including the Corpus of Contemporary American English (“COCA”).³⁸

Just this year, corpus linguistic analysis arose twice in the Sixth Circuit. Judge Thapar issued a concurring opinion that relied on his corpus linguistic analysis,³⁹ 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.⁴⁰

The growing use of dictionaries and corpus linguistics in interpretation is likely to continue. The Supreme Court’s next term includes a number of blockbuster interpretation cases, including whether the Eighth and Fourteenth Amendments permit states to abolish the insanity defense,⁴¹ and whether Title VII’s prohibition of discrimination “because of . . . sex”⁴² prohibits discrimination against employs for being lesbian, gay, bisexual, or transgender.⁴³ Corpus linguistics scholars have already filed a brief in the latter case.⁴⁴

Despite the enthusiasm surrounding dictionaries and corpus linguistics, there is surprisingly little work assessing what these tools actually do in legal interpretation. When one employs dictionaries or corpus linguistics in interpretation, these are properly understood as *empirical tools*, which might be inaccurate. Although the use of dictionaries and corpus linguistics seems to grow more sophisticated,⁴⁵ their accuracy has never been rigorously assessed.⁴⁶ There are important critiques of these methods from an external theoretical perspectives,⁴⁷ but we might also take an internal perspective, considering whether these methods even succeed on their own terms. Theories relying on these tools typically *assume* that dictionaries and corpus linguistics are accurate reflections of ordinary meaning,⁴⁸ “the ordinary user of the English language,” but the critical question remains: Is that assumption true?

This Article develops a novel method within “experimental jurisprudence” to test the accuracy of dictionaries and corpus linguistics in interpretation. This provides evidence about the justifiability of core methodologies of theories that rely on these tools. Insofar as a legal interpretive theory relies upon dictionary definitions or patterns of word usage,

³⁶ Lee & Mouritsen, *supra* note 13; *see also* Calhoun *supra*; 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. (2017); Evan Zoldan, *Corpus Linguistics and the Impossible Dream of Objectivity* (draft manuscript).

³⁷ *Muscarello v. United States*, 524 U.S. 125, 129 (1998).

³⁸ *E.g.* *People v. Hyarris*, 885 N.W.2d 832, 838 (Mich. 2016).

³⁹ *Wilson v. Safelite Group, Inc.* No. 2:16-cv-00877 (6th Cir. 2019).

⁴⁰ *William Wright v. Stephen Spaulding*, No. 17-4257 (2019).

⁴¹ *Kahler v. Kansas* (Kan. 2019).

⁴² 42 U.S.C. § 2000e-2(a)(1).

⁴³ *Bostock v. Georgia* (2019).

⁴⁴ Brief for Amici Curiae Corpus-Linguistics Scholars Professors Brian Slocum, Stefan Th. Gries, and Lawrence Solan in Support of Employees, *Bostock v. Georgia* (2019).

⁴⁵ *E.g.* Lee & Mouritsen, *supra* note 13; Lawrence M. Solan & Tammy Gales, *Corpus Linguistics as a Tool in Legal Interpretation* BYU L. REV. (2018).

⁴⁶ For other important empirical studies of judicial interpretation, see 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); 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. Donald L. Drakeman, *What’s the Point of Originalism*, 37 HARV. J. L. & PUB. POL’Y 1124; Jamal Green, 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.

⁴⁷ *E.g.* Brian G. Slocum, *Ordinary Meaning and Empiricism*, 40 STATUTE L. REV. 13 (2019); Victoria Nourse, *Picking and Choosing Text: Lessons for Statutory Interpretation from the Philosophy of Language*, 69 FLA. L. REV. 1409 (2017).

⁴⁸ For a demonstration of the use of corpus linguistics see Lee & Mouritsen, *supra* note 13. However, that demonstration is not necessarily a demonstration of a *reliable* method of corpus linguistics.

the study here also provides evidence about the success of such legal theories. Part I outlines the background to these debates and legal theories on 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 whether these tools are accurate measures of 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. How can we evaluate the accuracy of a 18th century dictionary or even 1980s corpus linguistics search without facts about the way in which the contested term was actually understood at the time (i.e. without the best data about its original ordinary meaning)? However, what we can do is evaluate whether *modern* uses of contemporary dictionaries and corpus linguistics reflect terms' *modern* meanings. If a tool (e.g. dictionary-use) performs poorly in modern interpretation, so long as there are no historically distinguishing factors (i.e. reasons that use of an 18th century dictionary is *more* accurate in historical interpretation than use of a modern dictionary is in modern interpretation), this gives us some evidence that the method is also unreliable in estimating original meaning.

Part III begins with a philosophical thought experiment: Do dictionaries and corpus linguistics accurately reflect ordinary meaning? Reflection on some definitions and patterns of word usage raises doubts.

Part IV presents a robust experimental investigation of dictionaries and corpus linguistics, testing their reliability and revealing some of the cognitive processes underlying their uses. Experimental studies (total N = 4,162) show systematic divergences among the verdicts delivered by ordinary concept use, dictionary use, and corpus linguistics use. For example, the way in which people today apply the concept of a vehicle is systematically different from the way in which people apply the modern dictionary definition of a “vehicle” or modern corpus linguistics data concerning vehicles. This pattern of results arises across samples of ordinary people (N = 3,834), “elite-university” law students (N = 230), and United States judges (N = 98).

Section IV.E contains a crucial summary and interpretation of the main experimental findings. There are ten noteworthy findings. Three concern the surprising similarity between expert and non-expert interpreters (e.g. judges vs. the general population). Judges, law students, and ordinary people were strikingly similar in their ordinary conceptual judgments (IV.E.1), use of dictionaries (IV.E.3), and use of corpus linguistics (IV.E.4).

Another important finding concerns the modern ordinary meaning of various terms; in *many* cases, there was significant disagreement about ordinary meaning—for judges, law students, and the general population alike (IV.E.2). That is, although there was consistency across these groups, *within* every group there was disagreement. For example, ordinary people were divided on whether a canoe is a vehicle; the same is true for law students and judges.

Two more findings concern the process of using dictionaries and corpus linguistics. Across the studies, users of corpus linguistics tended to identify prototypical examples (e.g. a car is a vehicle) better than non-prototypical examples (e.g. a moped or airplane is a vehicle) (IV.E.5). Conversely, users of dictionaries sometimes made more extensive judgments (e.g. a pair of rollerskates is a vehicle) (IV.E.6).

The next findings are that the verdicts of both corpus linguistics and dictionary use diverged from ordinary meaning, in a large number of cases (IV.E.7). Moreover, the verdicts of corpus linguistics and dictionary use diverged from each other (IV.E.8).

The final two sub-sections of Section IV.E quantify the rate at which the judgments of groups using dictionaries or corpus linguistics diverged from ordinary understanding of the term or phrase. Section IV.E.9 considers the results across a range of plausible assumptions. Across all these different models, each using reasonably conservative estimates, dictionaries and corpus linguistics had between 20-35% divergence rates or “error rates.” 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 corpus linguistics data would lead users to the wrong judgment fairly often: once in every three to five cases.

Section IV.E.10 notes that this 20-35% *average* rate of error may not tell the whole story. The experiments included a number of relatively easy categorizations (e.g. whether a car is a vehicle). Insofar as real legal decisions often concern comparatively harder categorizations (e.g. whether an airplane is a vehicle), it may also be instructive to consider the maximum error rate: What percent of (e.g.) judges using dictionary or corpus linguistics evaluated *the hardest interpretive question* incorrectly? Across all levels of expertise, the data suggest that in some examples, relying on a dictionary definition or corpus linguistics data led 80-100% of users to the incorrect verdict.

Part V considers whether the processes discovered by the experiments arise commonly in practice: Do legal uses of dictionaries tend to reflect broad, extensive interpretations, while legal uses of corpus linguistics tend to reflect narrow, prototypical uses? Part V.A finds that caselaw tends to refer to dictionary definitions as “broad” more often than as “narrow,” although dictionary definitions are sometimes narrowed by emphasizing context, which definition is most relevant, or particular features of the definition(s). Corpus linguistics has not yet been used frequently in caselaw, but many uses recommend a narrow meaning.

Nevertheless, this dictionary-extensive, corpus linguistics-prototypical relationship does not tell the entire story. Part V.B considers whether the nature of dictionary definitions—they generally recommend broad extensive meanings, but ones that can be narrowed—may also admit of motivated political decision making. The Part considers two examples from the Bill of Rights, from the Second and Eighth Amendments, each of which contains three terms: “keep and bear arms” and “cruel and unusual punishment.” Republican-appointed justices, at the Supreme Court and Circuit Court level, more frequently cite dictionaries to interpret the former set of terms broadly. Conversely, in Eighth Amendment cases, Republican-appointed justices interpret the broad dictionary definition narrowly. Although Democratic-appointed justices cite dictionaries less frequently, when they do, the pattern is reversed: dictionaries indicate that “cruel and unusual punishment” is broad, but that “keep and bear arms” is narrow. The dictionary-extensive, corpus-narrow relationship holds for other divisive examples. “Emoluments” seems broad when scholars consider dictionaries, but narrow when they consider corpora. So too for “Commerce.” The experimental findings show what consulting dictionaries and corpus linguistics typically provide, in the absence of other interpretive motivations and also in the absence of context. Section V.B is consistent with a large body of extant empirical literature showing that motivated reasoning may also play a role in interpretation.

Part VI elaborates the implications of the experimental results. Constructively, the results provide guidance for theories of interpretation. The results demonstrate one systematic pattern of judgment; while corpus linguistics tends to track prototypical uses, dictionaries tend to elicit more extensive ones. For example, consider a non-prototypical

vehicle, like an airplane. Today, people generally judge that an airplane is a vehicle.⁴⁹ That judgment is also reflected by participants' use of modern dictionaries, but users of corpus linguistics are hesitant to categorize airplanes as vehicles.⁵⁰ A similar pattern arises for other category members; corpus linguistics sometimes provides good evidence about prototypical members, but often fails to provide good evidence about non-prototypical members. Insofar as legal texts indicate the relevance of a very extensive concept application, dictionaries may be more useful. But insofar as legal texts indicate the relevance of only highly prototypical uses, corpus linguistics is more helpful.

Critically, the results support common fallacies in the use of dictionaries and corpus linguistics. For example, consider the "Non-Appearance Fallacy," the (false) claim that absence of a usage from a large corpus indicates that the usage is not part of the ordinary meaning. For example, 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 "vehicle," 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 (e.g. in a corpus).

More broadly, 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 corpus linguistics—two central tools of discovering ordinary meaning—are unreliable in interpretation, with error rates plausibly in the range of 20-35% and perhaps as high as 80%. This shifts the argumentative burden to those who rely on these tools to provide a principled account of the use of these tools and a demonstration of how error can be avoided.

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 prototypical criterion indicates that these entities are not vehicles. The findings show that dictionaries or corpus linguistics sometimes track *one* of these criteria—often dictionaries track the extensive criterion and corpus linguistics tracks the prototypical one—but a hard legal-philosophical question remains: Which of these *should* serve as a criterion in legal interpretation? Insofar as good reasons underlie both criteria, the results suggest that dictionary definitions, 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, legal interpreters will likely have to look beyond the dictionary and corpus—to the legal text's context, history, and purpose.

⁴⁹ See Part IV, *infra*.

⁵⁰ *Id.*

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 ironically unclear.⁵¹ Many treat “ordinary” and “plain” meaning synonymously.⁵² And recent work on corpus linguistics in originalist and textualist interpretation treats “ordinary” and “public” meaning coextensively when considering non-specialized terms and phrases.⁵³

Although there may be important differences among these various terms, 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 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 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.⁵⁴

This Article does not take a position regarding whether ordinary meaning *should* serve as a criterion of legal interpretation in any domain. Rather, the Article develops implications for various competing theories of legal interpretation.⁵⁵ That said, 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.”⁵⁶ Legal interpretation that reflects ordinary meaning also promotes reliance values, helping ordinary people manage their expectations and fostering coordination.⁵⁷ Moreover, interpreting a text in line with its ordinary meaning may prevent adjudicators from imposing their own personal beliefs in interpretation⁵⁸ and opportunistic behavior by the document’s drafters.⁵⁹ Empirical evidence suggests contracting parties might

⁵¹ 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 13; Richard A. Posner, *Statutory Interpretation—in the Classroom and in the Courtroom*, 50 U. CHI. L. REV. 800 (1983); *see also* SLOCUM, *supra* note 15; Fallon, *supra* note 15.

⁵² *See, e.g.* Stephen C. Mouritsen, *supra* note 14. Others draw a distinction such that “ordinary meaning” refers to a more de-contextualized meaning while “plain meaning” refers to the meaning the relevant term, phrase, or clause has *within* the contract, statute, or other legal document.

⁵³ *See* Lee & Mouritsen, *supra* note 13.

⁵⁴ Oliver Wendell Holmes, *The Theory of Legal Interpretation*, 12 HARV. L. REV. 417 (1899).

⁵⁵ *See* Part IV *infra*.

⁵⁶ James A. MacLeod, *Ordinary Causation: A Study in Experimental Statutory Interpretation*, 94 IND. L.J. 957, 979 (2019).

⁵⁷ *See* Gilson et al. *supra* note 17, at 40-41; Scott *supra* note 17.

⁵⁸ Uri Benoliel, *The Interpretation of Commercial Contracts: An Empirical Study*, 69 ALA. L. REV. 469, 475 (2017).

⁵⁹ CATHERINE MITCHELL, *INTERPRETATION OF CONTRACTS* 113 (2007).

generally prefer contracts to be interpreted by their ordinary meanings.⁶⁰ Further, interpreting laws in accord with their ordinary meaning promotes the publicity and ordinary clarity of law.⁶¹

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.⁶² Across diverse areas of legal interpretation—contracts, wills, trusts, deeds, patents, statutes, regulations, treaties, and constitutions⁶³—the ordinary meaning of the legal text is 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 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. 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 readers, which is not necessarily what the drafters aimed to accomplish in drafting the laws.⁶⁴

Of course, even in interpretation seeking to uncover intent, ordinary meaning is often taken to be the best 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,”⁶⁵ and most courts use plain meaning rules to establish this meaning, excluding extrinsic evidence when the will’s meaning is plain.⁶⁶ 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.

Sometimes interpreters seeks 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 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

⁶⁰ Benoliel, *supra*; Lisa Bernstein, *Private Commercial Law in the Cotton Industry: Creating Cooperation Through Rules, Norms and Institutions*, 99 MICH. L. REV. 1724, 1725 (2001); Lisa Bernstein, *Merchant Law in a Merchant Court: Rethinking the Code’s Search for Imminent Business Norms*, 144 U. PA. L. REV. 1765, 1769 (1996)..

⁶¹ See generally WILLIAM N. ESKRIDGE, JR., ABBE R. GLUCK & VICTORIA F. NOURSE, *STATUTES, REGULATIONS, AND INTERPRETATION: LEGISLATION AND ADMINISTRATION IN THE REPUBLIC OF STATUTES* (2014); Antonin Scalia, *Common-Law Courts in a Civil-Law System: The Role of United States Federal Courts in Interpreting the Constitution and Laws*, TANNER LECTURES 88 (1995).

⁶² Holmes, *supra* note 54; see also notes 1-12 *supra*.

⁶³ See notes 1-9 *supra*.

⁶⁴ See generally Solum, *supra* note 9; 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 (2016).

⁶⁵ RESTATEMENT (THIRD) OF PROPERTY: WILLS AND OTHER DONATIVE TRANSFERS §10.1 (2003).

⁶⁶ JESSE DUKEMINIER & ROBER TH. SITKOFF, *WILLS, TRUSTS, AND ESTATES* 328 (2013).

⁶⁷ Phillips, Ortner & Lee, *supra* note 64(citing 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 13, at 792. Lee & Mouritsen argue for this equivalence, except in the case of specialized legal language (e.g. “bill of attainder” or “parol evidence”). Part IV’s experiments and this paper’s arguments set aside study of specialized legal language. As such, following Lee & Mouritsen, I treat ordinary meaning to be equivalent to a text’s communicative content or public meaning.

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 meaning, like dictionaries or patterns of word usage across corpora.⁷²

B. Theories of Interpretation

Take the best-known hypothetical in legal interpretation: “no 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 (2019) 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, 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 start with the ordinary meaning of the term “vehicle.” 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.”⁷⁶

Textualist theories of statutory and constitutional interpretation would similarly look to the ordinary meaning of “vehicle” in the statute. Thirty years ago, Justice Scalia

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 *Id.*; MacLeod, *supra* note 56, at 985.

⁷³ H.L.A. Hart, *Positivism and the Separation of Law and Morals*, 71 *HARV. L. REV.* 593 (1958); *see also* Lee & Mouritsen, *supra* note 13, at 800.

⁷⁴ ESKRIDGE, JR. *supra* note 28.

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

⁷⁶ Eric A. Posner, *The Parol Evidence Rule, the Plain Meaning Rule, and the Principles of Contractual Interpretation*, 146 *U. PA. L. REV.* 533 (1998).

introduced the interpretive theory of “new textualism,”⁷⁷ noting that, “[e]very issue of law I resolve as 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—comes 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 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 might not look to the 2019 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 thereby constrains how it applies today. Looking to historical meaning will often make a difference, as ordinary meaning changes 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 or the statute’s purpose, or efficiency and consequences—it still plays a role as an important consideration in legal decision making 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 formalist and textualist theories of interpretation, and 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 clear language of

⁷⁷ *McBoyle v. United States*, 283 U.S. 25 (1931) (an airplane is not a vehicle, within the meaning of the criminal statute); *see also* *McReynolds v. Municipal Court of City of Ottuma*, 207 N.W.2d 792 (Iowa 1973) (an aircraft is a vehicle).

⁷⁸ Scalia, *supra* note 62.

⁷⁹ *See, e.g.*, Chiang & Solum, *supra* note 19 (on patents); Mouritsen, *supra* note 14 (on contracts).

⁸⁰ *See, e.g.*, Solum *supra* note 9; Lee & Mouritsen *supra* note 13.

⁸¹ *Id.*

⁸² *Id.*

⁸³ *E.g.* Stephen M. Griffin, *Pluralism in Constitutional Interpretation*, 72 TEX. L. REV. 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 Enterprises, LLC v. Marvel Characters, Inc.*, 716 F.3d 202, 313 (2d. Cir. 2013).

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.

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 arises in the well-known *McBoyle* case, in which Holmes’s interpretation of a criminal statute concluded that an airplane is not a vehicle in the ordinary sense of “everyday speech.”⁹¹ But it also arises in contract and insurance law, for example regarding whether a bicycle is a “vehicle” within the meaning of a plea agreement;⁹² and whether entities like watercrafts, airplanes, and motorcycles are “vehicles” within the meaning of insurance contracts.⁹³ 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 work demonstrates, the scope of the Article’s argument extends to the interpretation of many terms and phrases.⁹⁴

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, typically revealed by corpus linguistics methods.

A. Individual Intuition

A common source of interpretive evidence is individual intuition.⁹⁵ 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.⁹⁶ However, it is crucial to

⁸⁶Parks Real Estate Purchasing Grp. v. St. Paul Fire & Marine Ins. Co., 472 F.3d 33, 42 (2d Cir. 2006).

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

⁸⁸Lacks v. Fidelity & Cas. Co. of New York, 306 N.Y. 357, 363 (N.Y. 1954).

⁸⁹Peter Linzer, *The Comfort of Certainty; Plain Meaning and the Parol Evidence Rule*, 71 FORD. L. REV. 799, 800 (citing 2 MARGARET N. KNIFFIN, CORBIN ON CONTRACTS 2 s. 24.7, at 34). This is despite rejection of the plain meaning rule by both the Uniform Commercial Code and Restatement (Second) § 212 cmt. b. *Id.*

⁹⁰Mouritsen, *supra*.

⁹¹ *McBoyle v. United States*, 283 U.S. 25 (1931) (an airplane is not a vehicle, within the meaning of the criminal statute); *see also* *McReynolds v. Municipal Court of City of Ottuma*, 207 N.W.2d 792 (Iowa 1973) (an aircraft is a vehicle).

⁹² *Anthony v. State*, 329 P.3d 1027 (2014).

⁹³ *GEICO Marine Insurance Company v. Great Northern Insurance Company*, 2017 WL 4286394 (2017); *Certain British Underwriters at Lloyds of London v. Jet Charter Service*, 789 F.2d 1534 (1986); *Marasco v. Hopewell*, 2004 WL 2895973 (2004).

⁹⁴ 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* Part IV *infra*.

⁹⁵ *See generally* Solum, *supra* note 32, at 281-85.

⁹⁶ One reason to reject the constitution claim is that individuals sometimes make performance errors.

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 some easy circumstances, linguistic intuition is a helpful source in identifying ordinary meaning. Consider again the well-known example of local ordinance prohibiting “vehicles” from entering a park.⁹⁷ 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 (e.g. is a skateboard a vehicle?), linguistic intuition provides straightforward guidance in many other cases (e.g. baby shoulder-carriers may be brought into the park, cars must be kept out).

However in harder circumstances, linguistic intuition may not helpfully identify ordinary meaning—particularly in cases in which 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 (falsely) that they are good measures of the population’s consensus.⁹⁸ This concern is amplified by the fact that it is not always clear how an individual 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 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.”⁹⁹ Consider 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.¹⁰⁰ 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.”¹⁰¹

Individual intuition is recognized—by originalists and non-originalists alike—as an imperfect source of evidence in modern interpretation and a highly dubious source of evidence in historical interpretation. It is also a source of evidence whose errors are likely hard to identify in practice. Linguistic intuition often *feels* very compelling: Surely “dollar” 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, a method requires some guiding principles.

⁹⁷ This is discussed in much greater detail as the leading example in Parts III-V.

⁹⁸ See Solan & Gales, *supra* note 45, at 1333; Lawrence Solan, Terri Rosenblatt & Daniel Osherson, *False Consensus Bias in Contract Interpretation*, 108 COLUM. L. REV. 1268 (2008).

⁹⁹ Solum, *supra* note 32, at 281.

¹⁰⁰ This assumption also arises in some legal scholarship. See, e.g., Note, *The Twenty Dollars Clause*, 118 HARV. L. REV. 1665 (2005).

¹⁰¹ Solum, *supra* note 32, at 282.

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 Lee & Mouritsen note, the Webster’s Third New International Dictionary (1961) defines vehicle as a “carrier” or “agent of transmission.”¹⁰²

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.¹⁰³ 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.

In the search for historical meaning, another important limitation is that historical dictionaries are less frequent than modern ones. Two important English language dictionaries are published in 1755 and 1828.¹⁰⁴ 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 drafted in 1789? Relatedly, historical dictionaries were often the product of a single person’s efforts.¹⁰⁵ 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.¹⁰⁶ And their use is not without some initial plausibility: It is reasonable to suspect that the ordinary meaning of a term is often reflected well by its definition. Whether this suspicion is true is an open empirical question.

C. Linguistic Usage Data via Corpus Linguistics

A final source of interpretive evidence is corpus linguistics. Corpora are sets of language data, containing text from books, newspaper articles, online publications, and other sources.¹⁰⁷ In recent years, legal corpus linguistics has evolved from smaller searching to a “big data” approach.¹⁰⁸

¹⁰² Lee & Mouritsen, *supra* note 13, at 800.

¹⁰³ See James Brudney & Lawrence Baum, *Oasis or Mirage: The Supreme Court’s Thirst for Dictionaries in the Rehnquist and Roberts Era*, 55 WM. & MARY L. REV. 483 (2013); *see also* Solan & Gales, *supra* note 45, at 1334.

¹⁰⁴ Solum, *supra* note 32, at 283.

¹⁰⁵ *Id.*

¹⁰⁶ Calhoun, *supra* note 35.

¹⁰⁷ *See generally* DOUGLAS BIBER, SUSAN CONRAD & RANDI REPPEN, CORPUS LINGUISTICS: INVESTIGATING LANGUAGE STRUCTURE AND USE (1998).

¹⁰⁸ *See, e.g.,* Stefan Th. Gries & Brian G. Slocum, *Ordinary Meaning and Corpus Linguistics*, 6 B.Y.U. L. REV. (2017); Neil Goldfarb, *A Lawyer’s Introduction to Meaning in the Framework of Corpus Linguistics*, 6 B.Y.U. L. REV. (2017); Carissa Byrne Hessick, *Corpus Linguistics and the Criminal Law*, 4 B.Y.U. L. REV. (2018); Solan & Gales, *supra* note 45.

The most prominent defense of this “new corpus linguistics” approach is Lee & Mouritsen’s *Judging Ordinary Meaning*.¹⁰⁹ That paper advocates a promising account of corpus linguistics use in determining the ordinary meaning of legal texts, which can provide an objective, scientific, data-driven approach to legal interpretation.

The core of Lee and Mouritsen’s analysis contains two types of corpus searches: “collocation” and “keywords in context.” Collocation searches in a corpus show the words that are most likely to appear in the same context as the search term. A “keywords in context” search presents the user of corpus linguistics with example of the term in context.

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 co-locates 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.”¹¹⁰ For example, in a modern search, the top collocates of “vehicle” include “electric,” “motor,” “gas,” “autonomous,” and so on. As Lee and Mouritsen infer, “the collocates of *vehicles* . . . strongly indicate *automobiles* as a likely candidate for the most common use of the term.”¹¹¹

Next they conduct a keywords in context search. This 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.”¹¹²

It is important to note that this popularized use of corpus linguistics—focused on collocation and examples in context—is very different from versions of corpus linguistics that attempt to build more complex statistical or computation models of meaning, or those that use algorithmic processes (e.g. using word2vec to analyze word embeddings in a multi-dimensional vector space). The reason for this choice is likely motivated by practicality concerns. This simpler corpus method—in which a human evaluates patterns of language use across a range of written sources—is one that many legal interpreters (e.g. judges) can employ cheaply and swiftly. A 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.

III. ORDINARY MEANING THEORY’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 central critique concerns accuracy: does the use of a dictionary lead interpreters to accurate verdicts regarding ordinary meaning? 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 III.A. articulates these important critiques. The Article’s focus remains on the central *accuracy* critique, as well as the inconsistency

¹⁰⁹ *Supra* note 133.

¹¹⁰ Lee & Mouritsen, *supra* note 13.

¹¹¹ *Id.*

¹¹² *Id.*

concerns, but various other critiques remain relevant when assessing dictionary and corpus linguistics use.¹¹³

With these empirical questions in mind, Section III.B. turns a critical eye to the use of dictionaries and corpus linguistics in interpretation. It poses a thought experiment: What might future interpreters glean from modern dictionaries and corpus linguistics in interpreting a modern legal text? This sheds light on what dictionaries and corpus linguistics might commonly suggest to legal interpreters about ordinary meaning. Of course, this thought experiment is merely suggestive. In Part IV I present 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 the critique leveled at one source of evidence (e.g. individual intuition, dictionary-use, or corpus linguistics). Most of these are “internal” empirical critiques. That is, they do not criticize interpretive theories 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: Does the task’s *own* empirical assumptions withstand scrutiny?

<i>Critique</i>	<i>Example of Claim</i>
Accuracy: The method recommends a false verdict concerning ordinary meaning	Some method (e.g. dictionary-use) recommends an interpretive result that is inconsistent with the truth concerning ordinary meaning.
Inconsistency among interpreters: Different people using the method reach different verdicts	Different judges have different linguistic intuitions about the 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 sufficiently similar 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 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 corpus linguistics, as it is currently used in ordinary meaning interpretation.

¹¹³ Moreover, the empirical results of Part IV might also be taken to support various of these other critiques. For now, much of that analysis falls outside the scope of the present Article, which deals primarily with the central accuracy critique and the inconsistency critiques.

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.

Table 1. Empirical Critiques of Originalist Methods

These critiques can each be posed for different methods: individual intuition, dictionary-use, and corpus linguistics. This paper focuses primarily on the Accuracy critique 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 Corpus Linguistics Reliable? A Thought Experiment

With the increasing use of both dictionaries and corpus linguistics in interpretation, a crucial question looms: Are these methods actually achieving their aims? Often, interpreters simply assume that dictionary-use and corpus linguistics reflect facts about ordinary meaning. 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 18th century dictionary reflects the ordinary meaning of the time, we need true facts about ordinary meaning in the relevant context of the 18th century. The fact that we lack access to any direct verification source motivates interpreters to use other methods (like dictionaries and corpus linguistics) to provide evidence about ordinary meaning.

But perhaps a Rosetta Stone of ordinary meaning is not the only option. Although we lack precise verification about historical (e.g. 1787) ordinary meaning, we are much more confident in modern (i.e. 2019) ordinary meaning. Our familiarity with modern meaning can help assess sources of historical interpretive evidence. Consider a thought experiment:

Modern Amendment: Imagine that a modern Amendment stated that "vehicles" must be registered with the federal government. Two-hundred years later, in 2219, a

legal dispute erupts concerning the 2019 ordinary meaning or “original public meaning” of “vehicle.” Would consulting 2019 dictionaries and corpus linguistics provide precise and reliable evidence about the Amendment’s original meaning?

Following Lee & Mouritsen’s recent defense of originalist and textualist methodology, consider a typical dictionary definition of a vehicle: “an agent of transmission; a carrier.”¹¹⁴ An interpreter in 2219 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 aren’t vehicles.¹¹⁵ The Amendment’s 2019 ordinary meaning is not that roller-skates should be registered. But relying on the dictionary would suggest precisely the opposite.

Corpus linguistics may fare no better. As Lee & Mouritsen note, the written word “vehicle” almost always refers to a car. And it most often appears near words associated with cars, like “electric” and “motor.”¹¹⁶ 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. 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 corpus linguistics. The worry is not that these methods get things wrong in some unusual or esoteric cases. Insofar as 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.¹¹⁷ This reflection should give interpreters pause when using dictionaries or 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 (e.g. dictionary-use and corpus linguistics) reflect original meaning. This assumption remains surprisingly underexplored,¹¹⁸ 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 this also provides evidence that the method is unreliable in historical (i.e. originalist) interpretation:

1. Empirical Claim: The modern use of a method (i.e. use of dictionaries or 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.

¹¹⁴ Lee & Mouritsen, *supra* note 13, at 800.

¹¹⁵ For empirical evidence of this claim see Part IV.

¹¹⁶ Lee & Mouritsen, *supra* note 13, at 847.

¹¹⁷ *E.g.* *McBoyle v. United States*, 283 U.S. 25 (1931); *McReynolds v. Municipal Court of City of Ottuma*, 207 N.W.2d 792 (Iowa 1973).

¹¹⁸ 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 Frank Cross suggests that originalism does not, in fact, effectively restrain willful judging. FRANK B. CROSS, *THE FAILED PROMISE OF ORIGINALISM* (2013). *See also* sources cited *supra* note 46.

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 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, *our* use of a modern tool is presumably at least as accurate in reflecting modern ordinary meaning than is *our* 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. Does our 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 CORPUS LINGUISTICS

This Part turns to an experimental test of the reliability of dictionaries and corpus linguistics in interpretation. To explore the reliability of these methods, this Section presents a series of experiments. There are two broad aims of the present studies. One is an aim that existed at the outset of the studies: to assess whether people's use of corpus linguistic and dictionaries reflect ordinary meaning. The second is an aim that arose during the course of running the studies. After the first experiment revealed that dictionaries and corpus linguistics sometimes provide very different verdicts about ordinary meaning, a natural question arose about what might explain those differences.

First, I consider whether the (modern) ordinary meaning of "vehicle" is reflected in people's application of data from (modern) dictionaries and corpora. The results indicate that corpus linguistics sometimes diverges from ordinary understanding and from the recommendation of dictionary definitions.

Next, I investigate the second question: Why do corpus linguistics judgments differ from dictionary judgments (and why do both sometimes differ from the verdicts supplied by ordinary concept use)? I draw on prototype theory to test the hypothesis that dictionaries tend to generate more extensive uses, while corpus linguistics data is associated with more prototypical examples. Prototype theory holds that concepts are associated with certain features, and category members that have more of those features are regarded as more central than category members that have less of those features.¹¹⁹ For example, a car is a prototypical vehicle. An airplane, though a vehicle, is not a prototypical vehicle.

One possibility is that corpus linguistics data might be seen as supplying the most useful information about *prototypical* category members. Corpus data provides details

¹¹⁹ See Eleanor Rosch, *Cognitive Representation of Semantic Categories*, 104 J. EXPERIMENTAL PSYCHOLOGY 192 (1975).

about the most frequent uses of a term and the most common words associated with the term. For example, corpus linguistics data about vehicles indicates that certain words often appear near “vehicle,” such as “motor” and “electric.” In other words, perhaps corpus data is really supplying the most helpful information about *prototypical* category members. I hypothesized that corpus linguistics is more helpful in identifying prototypical category members, while dictionaries tend to generate a more extensive sense of category membership. This is precisely what the experiments find.

The third series of experiments test an “expertise” objection to the findings about the unreliability of corpora and dictionaries. Are “elite-university” law students (e.g. at Harvard and Yale) or judges experts in the use of dictionaries or corpus data, such that those methods are more reliable in their hands? And does the process underlying the use of these tools differ among these groups? For these studies I used a larger range of questions, in light of the previous experimental results. The data are strikingly similar to the earlier studies on non-experts, suggesting that whatever expertise judges and law students have, it does not provide more reliable use of these interpretive methods. Moreover, similar processes seem to guide all participants’ uses of dictionaries and corpus linguistics.

The final experiment tests whether the discovery of the error of dictionaries and corpus linguistics is limited to the example about vehicles. It attempts to replicate the earlier finding, using ten different examples: vehicle, carry, interpreter, labor, weapon, tangible object, animal, clothing food, and furniture. As a robustness check, the final experiment also varies some of the earlier experimental choices. For example, it uses a different corpus (the Corpus of Contemporary English, rather than the News on the Web Corpus); and it uses the first full definition of the term from Merriam Webster 2019. The data indicate that the results are robust: use of corpus linguistics and dictionary use carries significant error across *all* of these examples.

This paper includes a large number of experimental studies. To avoid redundancy and improve clarity, only the most significant are discussed in the main text. The remainder can be found in Appendices, A, B, C, and D.

Section IV.E is a crucial summary and interpretation of the main experimental findings. Four concern the surprising similarity between expert and non-expert interpreters (e.g. judges vs. the lay population recruited online). Judges, law students, and ordinary people were strikingly similar in their ordinary conceptual judgments (IV.E.1; IV.E.2), use of dictionaries (IV.E.3), and use of corpus linguistics (IV.E.4).

Two more findings concern the process of using dictionaries and corpus linguistics. Across the studies, users of corpus linguistics tended to identify prototypical examples (e.g. a car is a vehicle) better than non-prototypical examples (e.g. a moped or airplane is a vehicle) (IV.E.5). Conversely, users of dictionaries sometimes made very extensive judgments (e.g. a pair of rollerskates is a vehicle) (IV.E.6).

Finally, the verdicts of both corpus linguistics and dictionary use diverged from ordinary meaning, in a large number of cases (IV.E.7). Moreover, the verdicts of corpus linguistics and dictionary use diverged from each other (IV.E.8). Section IV.E.9 and IV.E.10 considers these results across a diverse range of plausible assumptions. Across all these different models, each using reasonably conservative estimates, dictionaries and corpus linguistics had between a 20-35% error rate on average, and an 80-100% error rate for some of the hardest examples.

A. Experiment 1

The first experiment tests the verdicts delivered by dictionary and corpus use, as compared to ordinary judgments. To minimize researcher degrees of freedom, I used the first test case mentioned by Lee & Mouritsen’s recent article endorsing the use of corpus linguistics in legal interpretation.¹²⁰ That case is the well-known “no vehicles in the park” example. I also used the exact corpus method used in that article, and the dictionary definition of “vehicle” that the article recommends.

The experiment divided participants into three groups: corpus, dictionary, and “ordinary concept” participants. The corpus and dictionary participants received corpus or dictionary information for 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 and six participants were recruited from Amazon’s Mechanical Turk (52% female, 48% male, 0% non-binary, mean age = 35.8). 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.¹²¹

Materials and Procedure. Participants were randomly divided into one of three conditions: Dictionary, Corpus, or Concept. In each condition, participants received some information about a term. Afterwards, they rated whether ten items fell under the category.

In the Concept condition, participants were simply asked to consider the noun “vehicle.” Then they were asked to categorize ten entities. For 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:¹²²

- 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.¹²³ 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

¹²⁰ Lee & Mouritsen, *supra* note 13.

¹²¹ See Adam J. Berinsky et al., *Evaluating Online Labor Markets for Experimental Research: Amazon.com’s Mechanical Turk*, 20 POL. ANALYSIS 351 (2012); Gabriele Paolacci et al., *Running Experiments on Amazon Mechanical Turk*, 5 JUDGMENT & DECISION MAKING 411 (2010). The service is understood to provide high-quality data. See Michael Buhrmester, Tracy Kwang & Samuel D. Gosling, *Amazon’s Mechanical Turk: A New Source of Inexpensive, Yet High-Quality Data?*, 6 PERS. PSYCHOL. SCI. 3 (2011).

¹²² Note that this dictionary definition mirrors the one suggested by Lee & Mouritsen, *supra* note 13, at 800 (“One attested sense of *vehicle* is the notion of a ‘carrier’ or ‘agent of transmission’” (citing WEBSTER’S THIRD NEW INTERNATIONAL DICTIONARY 2539 (1961))). See also <https://www.merriam-webster.com/dictionary/vehicle>.

¹²³ This is a common method in linguistics. See, e.g. Jean Berko, *The Child’s Learning of English Morphology* 14 WORD 150 (1958) (describing the “wug test”). Thanks to Larry Solan and Tammy Gales for this suggestion.

method of analysis (ordinary conceptual competence, dictionary definition, or corpus data).¹²⁴

Corpus 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, e.g. “electric ailac;” “ailac charging;” “drove the ailac;” etc.

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

Next Corpus participants saw further examples of the term in context.¹²⁵ This corpus data is *precisely* what the advocates of corpus linguistics recommend.¹²⁶ 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, Corpus, and Concept conditions. Figure 1 indicates the proportion responding “yes” for each entity in each condition.

¹²⁴ 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 (e.g. a “painter”) or one that is made up (e.g. 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.

¹²⁵ “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 ailacs 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...”.

¹²⁶ Lee & Mouritsen, *supra* note 13.

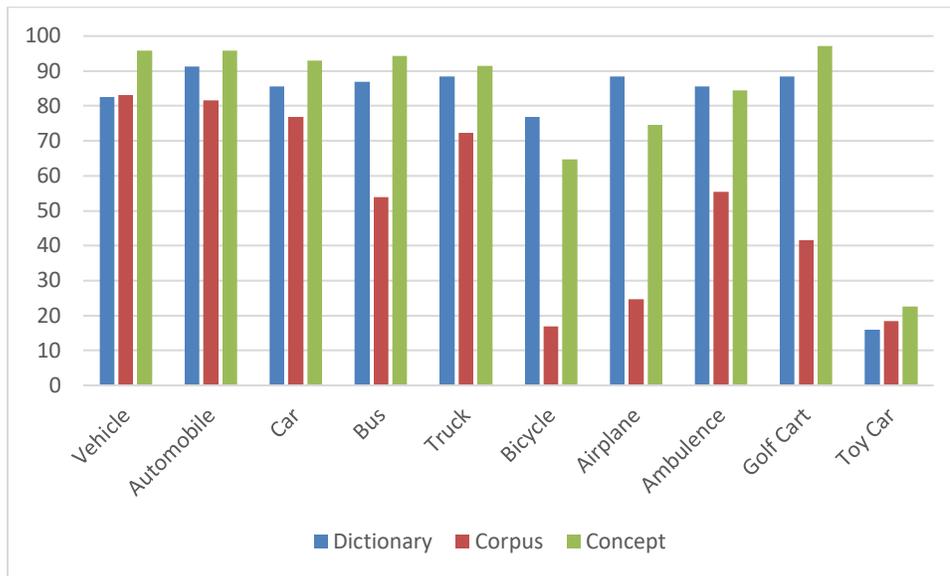


Figure 1. Percentage Responding Yes (the entity is a vehicle) by Dictionary, Corpus, Concept conditions

For each entity (e.g. car), I conducted a chi-square test to compare the proportion of yes responses across Dictionary, Corpus, and Concept conditions. See Figure A1 in the Appendix.

Notably, across the methods, some verdicts differ. Moreover, there is a systematic pattern: corpus fails to include a number of entities that are vehicles in the ordinary sense of the term. For example, consider entities like bicycle, airplane, and golf cart. Although there is some disagreement, most people classify these as vehicles. Yet users of corpus systematically judge that these are not vehicles.

As can be seen in Figure A1 (see Appendix), all differences are significant except for toy car. Conventionally, V of .1 is a “small effect,” .3 is a “medium effect,” and .5 is a “large effect.” By effect size conventions, there is a large effect for bicycle, airplane, and golf cart; a medium effect for bus and ambulance; a small effect for vehicle, automobile, car, truck; and no significant effect for toy car. The next natural question is what exactly explains each of these differences.

To address that question, I conducted pairwise chi-square tests to compare the proportion of “yes” responses for each item between conditions (Dictionary v. Corpus, Dictionary v. Concept, Corpus v. Concept). Notably, Dictionary and Corpus methods delivered different verdicts for five entities (bus, bicycle, airplane, ambulance, golf cart), and Corpus and Concept also diverged for those five.

Discussion

This first experiment represents a small test of the accuracy of corpus linguistics and dictionaries in reflecting ordinary meaning. Broadly speaking, dictionary-use was fairly consistent with people’s ordinary judgments: cars, busses, and trucks are vehicles, but a toy car is definitely not.

However, 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 corpus linguistics

made that categorization. The divergence was not limited to that example. For five of the ten entities, corpus was underinclusive.

B. Experiment 2: The Process of Using Dictionaries and Corpora

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

What explains these differences? One plausible hypothesis draws from research in linguistics and psychology on prototypes. According to prototype theory, people associate concepts with certain features, and “prototypical” category members are those that have most or all of those features. For example, both a robin and a penguin are birds, but a robin is a prototypical bird. Experimental studies have shown that people are faster in categorizing prototypical category members than non-prototypical ones. For example, people will 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, the more prototypical members are cited more frequently.¹²⁷ For example, if you ask someone to name a type of pet, “dog” would be cited more often than “kangaroo.”

I hypothesized that prototype theory might explain some of the differences between dictionaries, which report a broad definition, and corpus linguistics, which reports data indicative of the most popular uses. I hypothesized that corpus linguistics use helps identify prototypical examples, while dictionary use facilitates a more extensive representation. This experiment tests this hypothesis.

Method

Participants. One hundred and one participants were recruited from Amazon’s Mechanical Turk. Eighty-two passed a comprehension check question (51% female, 48.5% male, 1.2% non-binary, mean age = 35.8).

Materials and Procedure. Participants were trained to understand the difference between prototypical and non-prototypical category members (see Appendix A, Part III). Participants were then instructed to “Consider 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

Eighty-two participants answered all six check questions correctly and were included in the analysis. As predicted, there were significant differences between the prototypically and technically judgments across the ten entities, see Figure 2. Comparing

¹²⁷ See generally Eleanor Rosch, *Principles of Categorization*, in CONCEPTS: CORE READINGS 189 (1999).

these results to Experiment 1's results for Corpus and Dictionary participants reveals a striking similarity.

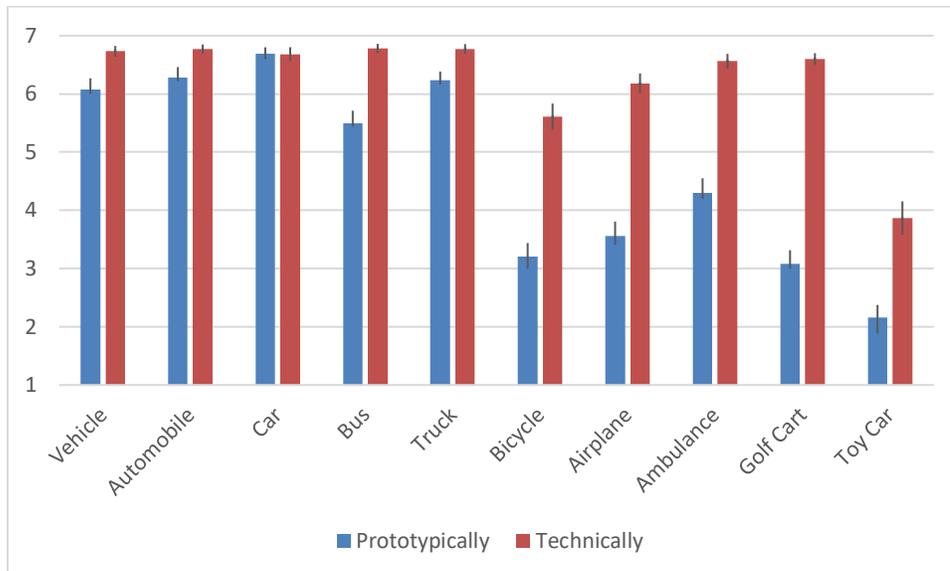


Figure 2. Mean ratings for “prototypal” and “technically” for ten entities. Error bars indicate standard error.

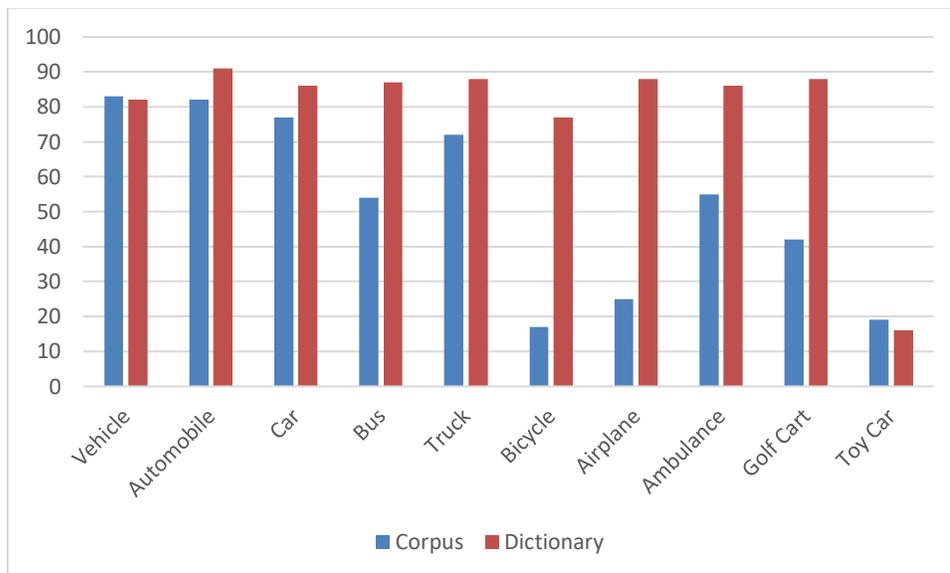


Figure 3. Percentage responding “yes” (entity is a vehicle) for ten entities by Corpus and Dictionary condition (Experiment 1).

To test the statistical significance of this relationship between Corpus-Prototypically and Dictionary-Technically, I conducted two tests for differences between correlations between (i) Corpus and Prototypically and Corpus and Technically, and (ii) Dictionary

and Prototypically and Dictionary and Technically.¹²⁸ In the first comparison, I considered the percentage of participants that rated each entity as a vehicle using the corpus, and correlated that with the ratings for prototypically and technically. Prototypically was significantly more correlated with Corpus, $z = 1.841$, $p = .0328$ (one-tailed). Technically was significantly more correlated with Dictionary, $z = 3.489$, $p = .0002$ (one-tailed).¹²⁹

	Corpus	Dictionary	Concept	Prototypically	Technically
Corpus	1				
Dictionary	0.51	1			
Concept	0.73	0.92	1		
Prototypically	0.93	0.58	0.72	1	
Technically	0.72	0.95	0.99	0.74	1

Figure 4. Correlation Matrix. Boxes indicate the relevant comparisons (Corpus is more correlated with Prototypically, and Dictionary is more correlated with Technically).

Discussion

Insofar as corpus linguistics elicits prototypical uses of a term but dictionaries elicit a more extensive sense, the former may be more appropriate in legal contexts calling for a prototypical uses and the latter more appropriate in contexts calling for more extensivist ones. For example, in the context of a rule “any and all vehicles are prohibited from the park,” one might reasonably think that the ordinary meaning of the rule bans even non-prototypical vehicles. But in the context of a rule “only cars, trucks, and other vehicles are prohibited from the park,” one might argue more persuasively that the rule bans only 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 (e.g. “vehicle”). But some might doubt whether this population contains the best users of dictionaries and corpus linguistics in legal interpretation. To appropriately test the reliability of corpus linguistics and dictionaries, one might argue, we should test legal experts who have the relevant background in interpretation.

¹²⁸ See Ihno A. Lee & Kristopher J. Preacher, *Calculation for the test of the difference between two dependent correlations with one variable in common* (2013), <http://quantpsy.org>.

¹²⁹ 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, $z = 3.489$, $p = .0005$ (two-tailed). Prototypically more correlated with corpus at a level short of the traditional cutoff for statistical significance, $z = 1.841$, $p = .0656$ (two-tailed). To match the previous experiments, this experiment uses only ten items (e.g. car, bus, bicycle, etc.). This limits the power of correlation analyses.

This objection is plausible, but it should not be taken to dismiss any significance of the results. After all, even if the previous results do not provide a strong inference into judges' cognition, they do provide good evidence about juror's cognition. And jurors, too, are statutory interpreters.¹³⁰

Nevertheless, this section addresses the "expertise" objection head-on. I tested a population of law students from Harvard, Yale, and Columbia and United States judges. There are a few reasons to think such experts might perform differently. For one, it may be that they have some training or expertise that enables them to use dictionaries or corpus linguistics in some expert way. Second, even if they do not have special expertise in 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 corpus, 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 includes 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 are entities including crutches, a baby should-carrier, and a zip-line.

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

Method

Participants. Approximately 700 professional judges were contacted by email to request voluntary participation in the study. Ninety-eight United States judges participated in an online experiment.¹³¹ Judges were recruited from state and federal courts and asked to categorize their years of experience (e.g. less than 1 year, 1-5 years, 6-10 years, etc.). Seventy-three judges reported their years of judging experience. Of those, 1.4% reported less than one year of experience, 17.8% reported 1-5 years of experience, 17.8% reported 6-10 years, 24.7% reported 11-15 years, 17.8% reported 16-20 years, 11.0% reported 21-25 years, and 8.22% reported over 26 years (1.37% reported "other").

Materials and Procedure. As in the previous experiments, participants were randomly assigned to either the Concept, 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 car, toy car. Participants immediately considered another set (presented in a randomized order): drone, skateboard, pair of rollerskates, "a non-functioning commemorative truck (e.g. a World War II Truck that has been decorated as a WWII monument)", baby stroller, electric wheelchair, horse-drawn carriage, wooden canoe, helicopter, moped, pair of crutches, pogo stick, baby shoulder-carrier, liferaft, and zip-line.

¹³⁰ Lawrence Solan, *Jurors as Statutory Interpreters*, 78 CHI-KENT L. REV. 1281 (2003).

¹³¹ $M_{age} = 59.3$, 33.8% female, 66.2% male, 0.0% non-binary.

Results

First, I conducted chi-square tests for differences among the three conditions. For twenty items, there was a significant difference. To further analyze these differences, I conducted follow-up pairwise chi-square tests. Again, there were a number of significant differences (see Appendix A).

Discussion

The results are strikingly similar to the results of non-experts. The judges' use of corpus linguistics and dictionary methods did not consistently reflect their ordinary judgments about category membership. For the full results and statistics, see Appendix A.

For many entities, the corpus linguistics judgment did not reflect that of judges' ordinary conceptual competence: consider bus, truck, airplane, ambulance, golf cart, horse-drawn carriage. For many others, the dictionary use did not reflect ordinary judgment: consider skateboard, roller-skates, WWII Truck, baby stroller-carrier, canoe, helicopter, 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, liferaft, and zip-line.

D. Experiment 4: Replication Across Ten Examples

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

This final experiment sought to test whether these findings replicate across different examples. To examine this question, this 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."¹³² The next three are inspired by important interpretation terms: "labor," "tangible object," and "weapon" (a modern version of "arms"). The final four are 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. For further detail, see Appendix D.

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 textualist interpretation,¹³³ this experiment simply used the first full definition of the relevant term, from Merriam-Webster 2019 Online.¹³⁴ In some cases, these definitions supplied some examples alongside the definition. For example,

¹³² See Lee & Mouritsen, *supra* note 13; see also Appendix D *infra*.

¹³³ See Lee & Mouritsen, *supra* note 13.

¹³⁴ Merriam-Webster, <https://www.merriam-webster.com/>.

“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.” 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. Two-thousand eight-hundred and thirty-five “general population” participants from the United States were recruited from Amazon’s Mechanical Turk ($M_{age} = 37.88$, 46.1% female, 53.6% male, .03% non-binary).

Materials and Procedure. The procedure was similar to 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). See Appendix D for full materials.

Results and Discussion

As a first analysis, I conducted chi-square tests to assess differences in categorizations among the ordinary concept, corpus, and full dictionary conditions. There were ten entities, each with twenty-five entities (and twenty-five comparisons). As such, I analyzed each example (e.g. vehicle; weapon) by assessing whether there was at least one significant difference, at a level corrected for twenty-five multiple comparisons.¹³⁵ If there was one significant difference at this level, this would suggest some unreliability among the methods for that example.

The results provide fairly strong evidence of a large degree of unreliability. For *all* ten examples (e.g. vehicle; tangible object), there were at least ten significant differences (out of twenty-five). For vehicle, there were eleven; for carry, there were twelve; for labor, fifteen; for interpreter, twenty; for tangible object, twenty-two; for weapon, fifteen; for furniture, eighteen; for animal, fourteen; for food, twenty; and for clothing, fourteen. See Appendix D for full results.

This is consistent with the earlier experimental results. For example, in Experiment 3, there were eighteen significant differences for judges (concerning vehicle) and twenty for law students (concerning vehicle).

As a second analysis, I conducted chi-square tests to assess differences in categorizations among the ordinary concept, corpus, and bare dictionary conditions. There were ten entities, each with twenty-five entities (and twenty-five comparisons). As such, I analyzed each example by assessing whether there was at least one significant difference, at a level corrected for twenty-five multiple comparisons.¹³⁶ If there was one significant difference at this level, this would suggest some unreliability among the methods for that example.

The results again provide strong evidence of a large degree of unreliability. For *all* ten examples, there were at least ten significant differences (out of twenty-five). For

¹³⁵ I compared the chi-square results to a Bonferonni-corrected p value of .002.

¹³⁶ I compared the chi-square results to a Bonferonni-corrected p value of .002.

vehicle, there were thirteen; for carry, there were eleven; for labor, twenty-two; for interpreter, thirteen; for tangible object, twenty-two; for weapon, fourteen, for furniture, seventeen; for animal, fifteen; for food, twenty; and for clothing, eleven. See Appendix D for full results.

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 non-experts are similar in (i) their ordinary judgments concerning common terms (e.g. “vehicle”), (ii) how they apply dictionary definitions, and (iii) how they apply corpus linguistics data about word usage.

Moreover, the results indicate that, perhaps surprisingly, “ordinary meaning” is not as clear as one might think. For a number of entity categorizations, participants were 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 corpus linguistics. Corpus linguistics categorizations are correlated with judgments of prototypicality, while dictionary categorizations are more extensive. Unsurprisingly, this implies that dictionaries and 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 Non-Experts Judge Meaning Similarly

First, consider the percentage of participants within each population responding that each entity is a vehicle.

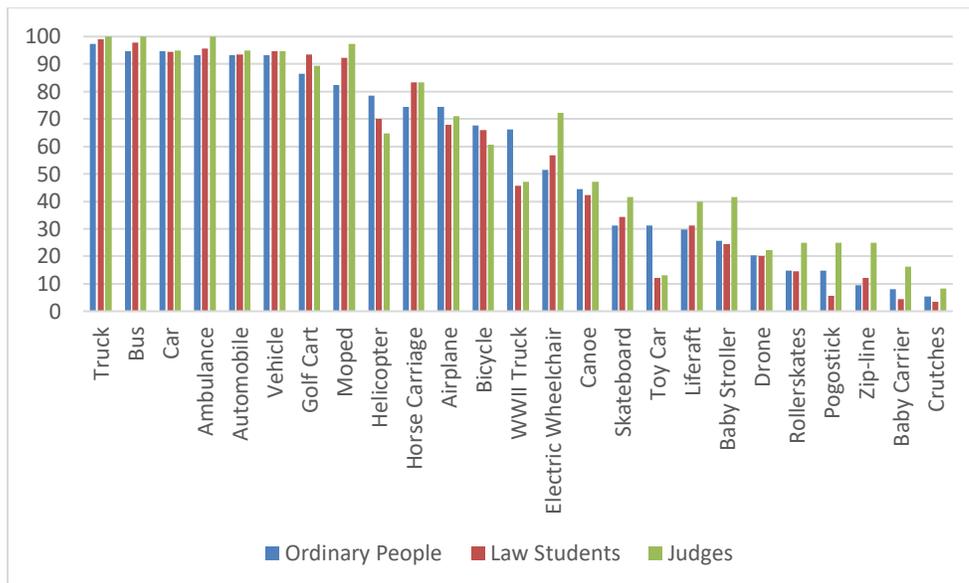


Figure 5. Percentage of participants responding “Yes” to “Is [entity] a vehicle?”

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*.

2. Ordinary Meaning is Often Unclear

Figure 5 indicates a second striking fact about ordinary concepts. 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 *facts* about ordinary meaning, but this result suggests that in some cases those facts 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 (in the same degree) across people with various levels of legal training and experience.

3. Judges and Non-Experts Apply Dictionaries Similarly

Next consider the percentage of participants within each population responding that each entity is a vehicle, according to the dictionary.

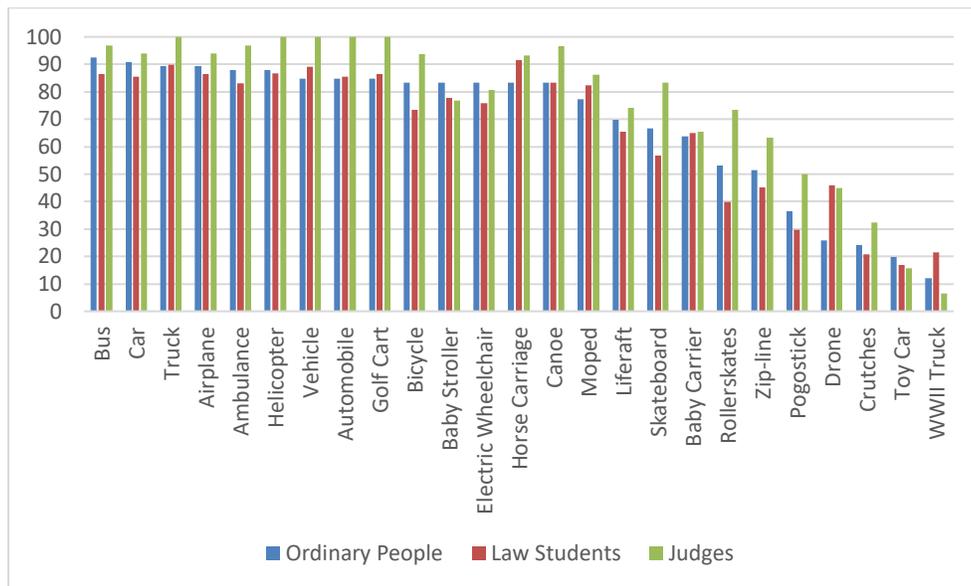


Figure 6. Percentage of participants affirmatively categorizing the entity according to the dictionary definition

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 MTurk participants do not disagree *as groups* about how to apply dictionaries—there is striking disagreement within groups for some

entities. Consider examples like the zip-line, pogostick, and drone. A substantial proportion of participants in every group categorized these as vehicles, while a substantial proportion did not.

4. Judges and Non-Experts Apply Corpus Linguistics Similarly

Finally, consider the percentage of participants within each population responding that each entity is a vehicle, according to corpus linguistics.

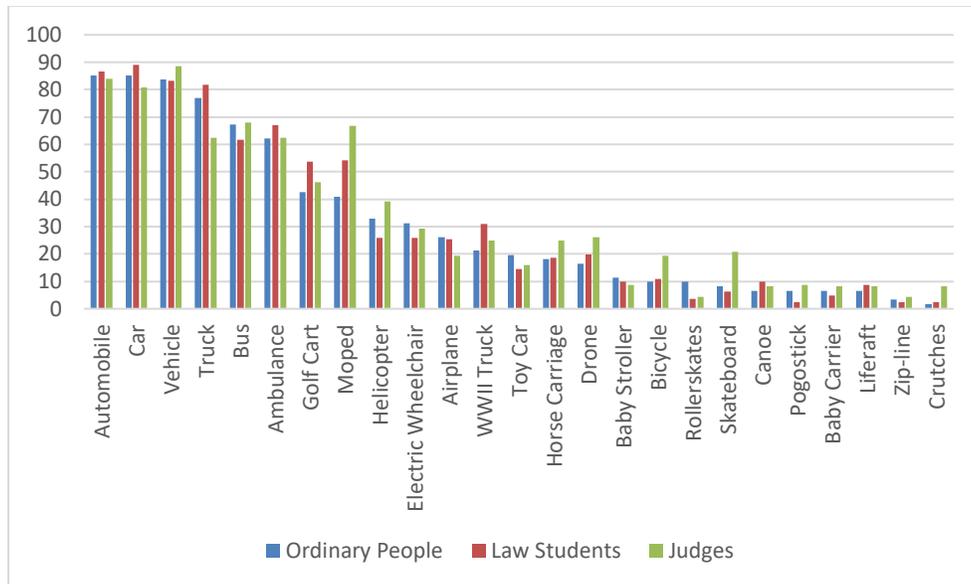


Figure 7. Percentage of participants affirmatively categorizing the entity according to the corpus linguistics data

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

Here again, we should note that although there is impressive agreement among the populations—judges, law students, and MTurk participants are not very different *as groups* in their applications of corpus linguistics— but 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. Corpus Linguistics Use Reflects Narrow, Prototypical Uses

Next consider the corpus and concept results for judges.

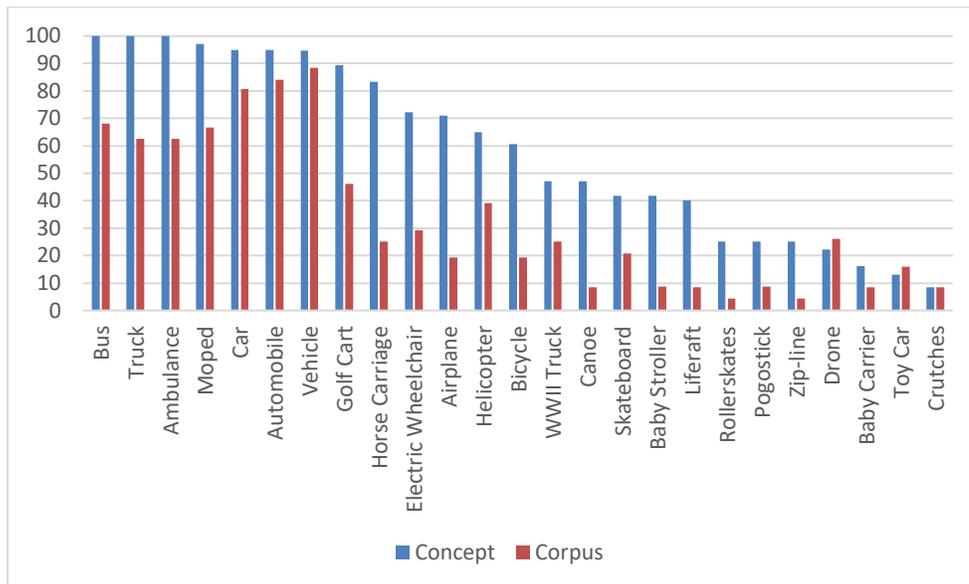


Figure 8. Comparison of percentage of judges responding “Yes” to “Is [entity] a vehicle?” to percentage of judges categorizing the entity as a vehicle according to corpus linguistics.

Although 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 corpus linguistics returned only a moderate endorsement of trucks as vehicles. Similarly, entities like a horse-drawn carriage, golf cart, airplane, helicopter, and bicycle are largely understood by judges as vehicles in ordinary language, but they are not classified as vehicle by judges using corpus linguistics.

Moreover, when we compare these proportions of judge’s categorizations to the rated prototypicality of the entities (as in Experiment 2), we find the same significant relationship between corpus linguistics and prototypicality (see Appendix A).

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 (e.g. canoe), dictionary-use tends to reflect that those *are* vehicles.

Moreover, when we compare these proportions of judge’s categorizations to the ratings of whether the entity is a “prototypical” vehicle or “technically” a vehicle (as in Experiment 2), we find the same significant relationship between dictionaries and the extensive question (see Appendix A). This suggests that unlike corpus linguistics, which tracks prototypical uses, dictionaries track a more extensive sense of meaning.

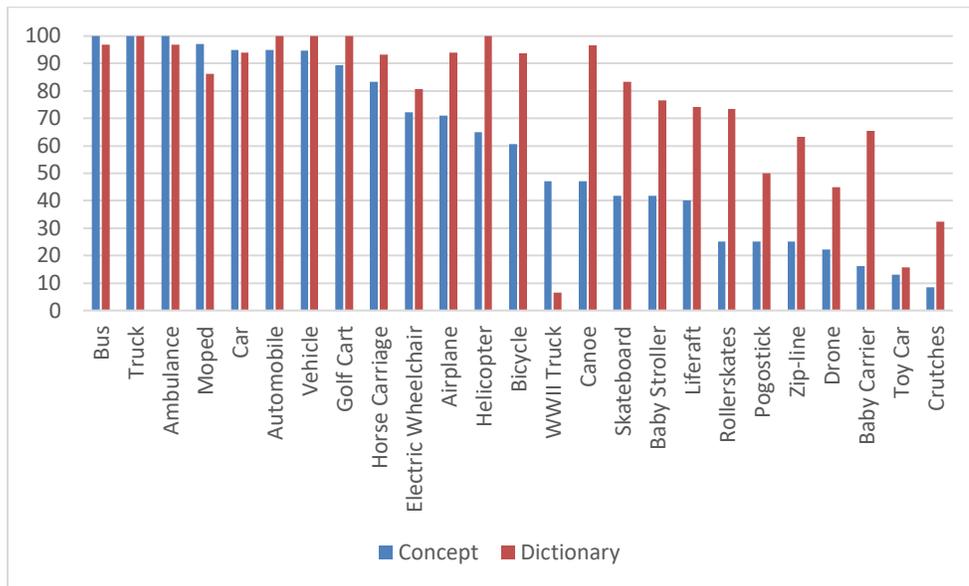


Figure 9. Comparison of percentage of judges responding “Yes” to “Is [entity] a vehicle?” to percentage of judges categorizing the entity as a vehicle according to dictionary use.

7. Dictionaries and 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 corpus linguistics. Insofar as these sources of evidence should be tracking the same “ordinary meaning,” this results suggests some serious risk of error in at least one of the methods.

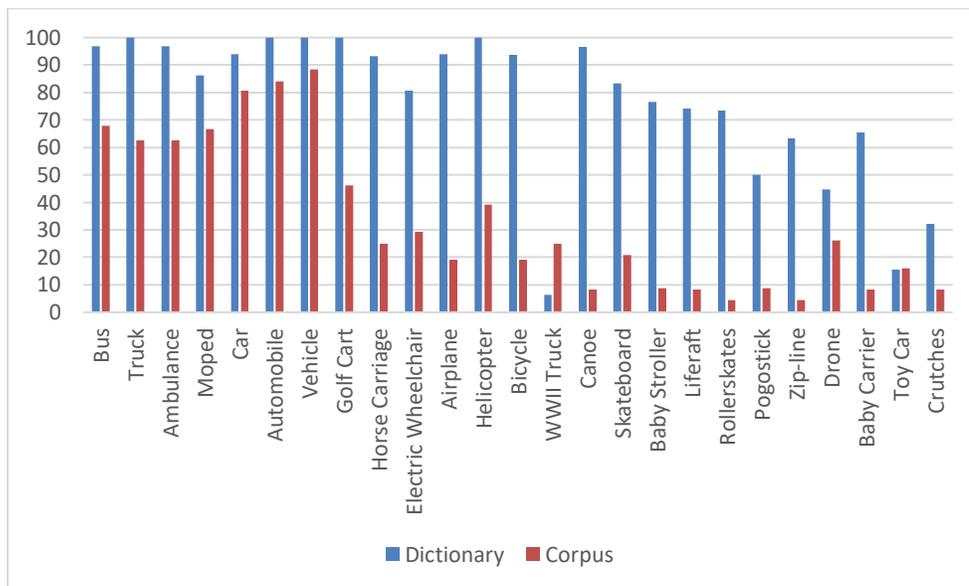


Figure 10. Comparison of percentage of judges categorizing the entity as a vehicle according to dictionary use and corpus linguistics use.

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

8. Dictionary and Corpus Linguistics Verdicts Diverge from Ordinary Judgment

Figures 9 and 10 indicate that, for many examples, use of dictionaries and corpus linguistics did not reflect ordinary judgments.¹³⁷ The most straightforward interpretation of this pattern of results is that dictionaries and corpus linguistics were not always accurate measures of modern meaning.

Considering just the results of Experiment 3, in some cases, corpus linguistics indicated that clear vehicles were not, in fact, vehicles. For example, busses, trucks, and ambulances were unanimously understood to be vehicles. Yet over one-third of those using corpus linguistics evaluated these as not vehicles. Conversely, in some cases the dictionary use indicated that clear non-vehicles were, in fact, vehicles. For example, dictionary-using judges overrated rollerskates and baby-shoulder carriers as vehicles, compared to judges' ordinary evaluation of those entities.

9. On Average, Corpus Linguistics and Dictionaries Had 20-35% Error Rates

This Section computes and considers the "error rates" for both dictionary and corpus linguistics methods, across Experiments 3 and 4. That is, it considers how often one relying upon a dictionary definition or corpus linguistics data would reach the *wrong* verdict about ordinary meaning.

To compute assess these error rates, we must make some assumptions about what percentage of agreement in the ordinary concept condition indicates that the use of part of 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 a simple "clear majority" of 50%.¹³⁸

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 (i.e. we need not commit that 50% is the right cutoff to determine 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% participants (in the ordinary concept condition) categorized something as a vehicle, then it counts as a 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 corpus linguistics made the same judgment. So, 32% of corpus users made a judgment (i.e. that a bus is not a vehicle) that is incorrect on these assumptions. On this assumption, there is a 32% error rate for the bus item for judges

¹³⁷ See also Appendix C *infra*; Appendix D *infra*.

¹³⁸ Ben-Shahar & Strahilevitz, *supra* note 71, at 1779.

using corpus linguistics. Repeating this process for all items (i.e. all 25 items in Experiment 3; and all 250 items in Experiment 4), we can compute an average error rate.

I performed these computations, using 50%, 75%, and 90% cutoffs, for the corpus 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.

A final 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 *very generous* to 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 corpus linguistics 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 Table 2 and 3, respectively.

Average 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 Population (Ten)	41.9%	39.1%	39.5%	25.1%

Table 2. Average error for corpus linguistic 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%	50.79%	22.4%
Law Students (Vehicle)	28.6%	35.9%	45.96%	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%

Table 3. Average error for dictionary judgments in Experiments 3 and 4, under different theoretical assumptions.

The important takeaway from these tables is that the error rates for dictionaries and corpus linguistics 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% to 35%. Sometimes it was higher (e.g. 50.8% for judges’ use of dictionaries, with a 90% cutoff); and sometimes it was lower (e.g. 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 20-35%) error rates suggest that one relying on dictionaries or corpus linguistics would reach the wrong verdict *in every three to five cases*.

Importantly, this “error rate” is not an estimate of how often users of dictionaries or 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 corpus linguistics data concerning a term*. As such, the error rate calculation is most significant for that specific type of—not uncommon—legal interpretation.¹³⁹

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 (e.g. whether a car is a vehicle; whether factory working is labor; whether a book is a tangible object). Insofar as real legal decisions concern comparatively more difficult categorizations (e.g. whether an airplane 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 (e.g.) judges using dictionary or corpus linguistics evaluated *the hardest interpretive question* incorrectly?

¹³⁹ See Nourse, *supra* note 26.

I conducted a similar analysis to that conducted in Section IV.E.9. But in this analysis, I computed the maximum error rate, under each of the different assumptions. As Tables 4 and 5 indicates, across all levels of expertise, the data suggest that in some examples, relying on a dictionary definition or corpus linguistics data led 80-100% of users to the incorrect verdict.

Maximum 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)	75.3%	73.1%	73.5%	37.1%

Table 4. Maximum error for corpus linguistic 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%

Table 5. Maximum error for dictionary judgments in Experiments 3 and 4, under different theoretical assumptions.

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

V. HOW JURISTS USE DICTIONARIES AND 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 legal uses of corpus linguistics tend to reflect narrow, prototypical uses? While corpus linguistics is relatively new in legal decision making, dictionaries are frequently cited. Section V.A surveys the pattern of citation and finds that caselaw tends to refer to dictionary definitions as “broad” significantly more often than as “narrow.” Moreover, while dictionaries are often understood to be broad, their definitions are sometimes narrowed by considering contextual features or which of multiple definitions is most relevant. Corpus linguistics has been used less frequently in caselaw, 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 other *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 corpus linguistics sheds light on these and other debates about ordinary meaning.

Although Section V.A suggests that this tendency of dictionary and corpus linguistics methods manifests in legal decision making, 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 corpus linguistics uses to adequately assess this claim, but Section V.B considers whether the property of dictionary definitions—they are generally broad but can be narrowed by some interpretive choices—may admit of political decision making. To examine this question the Part 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 punishment.” Republican-appointed justices, at the Supreme Court and Circuit Court level, more frequently cite Founding Era dictionaries to interpret terms like “keep” “bear” and “arms” broadly. Conversely, when Republican-appointed justices cite dictionaries in Eighth Amendment cases, the broad dictionary definitions are interpreted narrowly. Although Democratic-appointed justices cite dictionaries less frequently, when they 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 corpus linguistics, while there are also important limitations of the scope of that insight: Where dictionaries and corpus linguistics are cited, there are a number of other factors (e.g. a statute’s precedent or purpose; extrinsic evidence of contracting parties’ intentions) and biases (e.g. politically motivated reasoning) that add further complexity.

A. Dictionaries Often Supply “Broad” Senses and 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 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 rollerskates as vehicles, but most people do not consider those entities to be vehicles.

This result may seem less surprising if we reflect 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 an “agent of transmission” or a “carrier.” This definition is broader than the definition of the most prototypical vehicles. 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 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 citations to the terms “broad” and “narrow” within the same sentence (“/s”) or paragraph (“/p”) in the Supreme Court, and a sample from Lower Federal Courts and Circuit Courts.¹⁴⁰

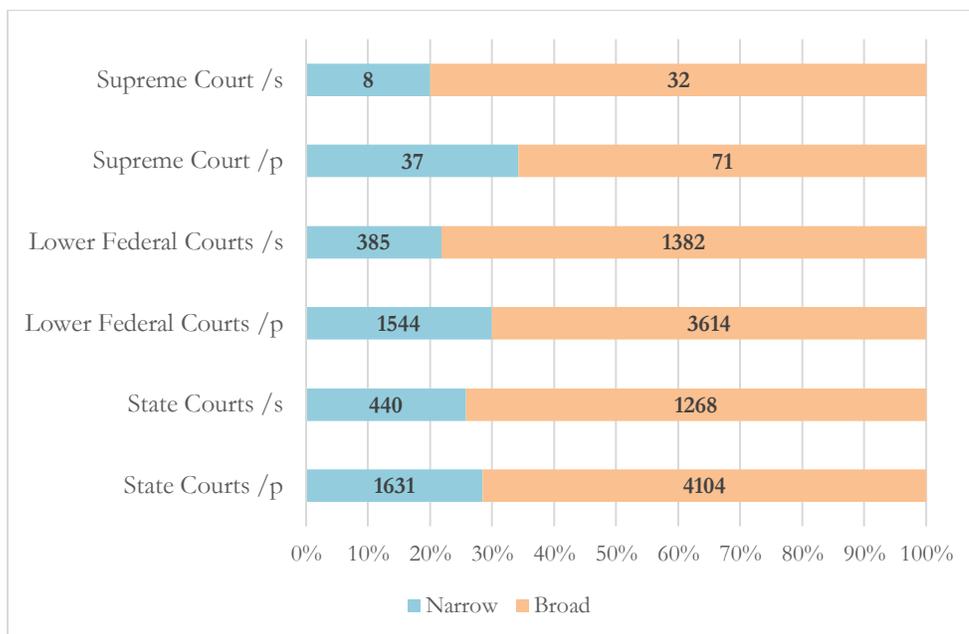


Figure 11. Court citations of “broad” and “narrow” in the same sentence (/s) or paragraph (/p) as “dictionary”¹⁴¹

This suggests that more often courts describe dictionaries as broad.¹⁴² 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

¹⁴⁰ For further detail, see notes 142-154 *infra*.

¹⁴¹ Searches conducted on Westlaw. “/s” indicates that the terms are within the same sentence; “/p” indicates that the terms are within the same paragraph.

¹⁴² Note, 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 is independent from the this claim that corpus linguistics provides evidence about whether dictionaries are typically described as broad or narrow.

uses,¹⁴³ Supreme Court within paragraph uses,¹⁴⁴ lower federal court within sentence uses,¹⁴⁵ lower federal court within paragraph uses,¹⁴⁶ state court within sentence uses,¹⁴⁷ and state court within paragraph uses.¹⁴⁸

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 was not statistically significant: Supreme Court within paragraph uses.¹⁴⁹ However for all other comparisons, the same pattern held: for Supreme Court within sentence uses,¹⁵⁰ lower federal court within sentence uses,¹⁵¹ lower federal court within paragraph uses,¹⁵² state court within sentence uses,¹⁵³ and state court within paragraph uses.¹⁵⁴ Overall, this pattern of results strongly suggests that dictionaries are more often cited in the context of “broad” than “narrow.”

To further evaluate the significance of this pattern, consider some of the Supreme Court “broad” dictionary examples:

- “That a dictionary definition is broad enough to encompass one sense of a word does not establish, however, that the word is ordinarily understood in that sense.”¹⁵⁵
- “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”¹⁵⁶
- “...the dictionary definitions of that word are very broad.”¹⁵⁷
- “Broad definitions of the term in modern and older dictionaries are unhelpful”¹⁵⁸

One striking feature of these uses is that, while most suggest dictionary definitions are broad, many cite this as a reason that legal outcome should *not* be constrained by the meaning recommended by the dictionary.

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:

- “To our knowledge *all* English dictionaries provided the narrow definition of ‘modify’ [connoting only moderate, and not fundamental, change]”¹⁵⁹

¹⁴³ Binomial $p = .0099$.

¹⁴⁴ Binomial $p = .0275$.

¹⁴⁵ Binomial $p < .0001$.

¹⁴⁶ Binomial $p < .0001$.

¹⁴⁷ Binomial $p < .0001$.

¹⁴⁸ Binomial $p < .0001$.

¹⁴⁹ Binomial $p = .2478$.

¹⁵⁰ Binomial $p = .0147$.

¹⁵¹ Binomial $p < .0001$. To provide a test case, I 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.

¹⁵² Binomial $p < .0001$.

¹⁵³ Binomial $p < .0001$. To provide a test case, I 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.

¹⁵⁴ Binomial $p < .0326$.

¹⁵⁵ *Taniguchi v. Kan Pacific Saipan, Ltd.*, 566 U.S. 560 (2012).

¹⁵⁶ *Yates v. U.S.*, 135 S.Ct. 1074, 1083 (2015).

¹⁵⁷ *Smith v. U.S.*, 508 U.S. 223 (1993).

¹⁵⁸ *Bullock v. BankChampaign, N.A.* 569 U.S. 267 (2013).

¹⁵⁹ *MCI Telecommunications Corp. v. American Tel. & Tel. Co.*, 512 U.S. 218 (1994).

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

- “Some [law dictionaries] define ‘firm’ [narrowly] But other dictionaries, while recognizing that narrow definition, also state that the word has a broader meaning...”¹⁶⁰
- “[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.’¹⁶¹
- “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, see Webster’s New International Dictionary 2406 (2d ed. 1934) (‘sovereignty,’ definition 3), but sovereignty in the narrow legal sense of the term, meaning a claim of right”¹⁶²

These empirical results are consistent with the experimental findings: Dictionary definitions often (but not always) supply a broad, extensive sense of meaning. Importantly, 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 context, (b) selecting a relatively narrower definition, where there are multiple definitions, or (c) emphasizing the necessity of particular features of the definition.¹⁶³

Corpus linguistics is relatively new and has yet to appear in a range of court decisions. So, inevitably, the conclusions that we can draw from caselaw practice are much more limited. 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 the Adoption of Baby E.Z.*,¹⁶⁴ Justice Lee analyzed the phrase “custody determination.” He considered 500 sample sentences from the Corpus of Contemporary American English, and reported that the most common use of “custody” was in the context of divorce, rather than in the context of adoption. 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 in a broader range of custody proceedings.¹⁶⁵ This 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*, corpus linguistics suggested that “discharge” was largely used to refer to a single shot of a firearm, rather than emptying the entire magazine.¹⁶⁶ This, too, is a less broad 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

¹⁶⁰ U.S. v. Cook, 384 U.S. 257, n.5. (1966).

¹⁶¹ Smiley v. Citibank (South Dakota) 517 U.S. 735, 746 (1996).

¹⁶² Boumediene v. Bush, 553 U.S. 723, 754, 128 S. Ct. 2229 (2008).

¹⁶³ Compare, for example, Breyer and Ginsburg’s opinions in *Muscarello v. United States*, 524 U.S. 125 (1998); and see also Scalia’s dissent in *Chisom v. Roemer*, 501 U.S. 380 (1991) (“representatives” does not include judges).

¹⁶⁴ 266 P.3d (Utah 2011).

¹⁶⁵ *Id.*

¹⁶⁶ 356 P.3d 1258 (Utah 2015) (Lee, J., concurring).

“expectations of privacy,” Thomas notes 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.”¹⁶⁷ This 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/exclusive recommendation concerning ordinary meaning. But the early judicial uses of corpus linguistics are suggestive of such a trend.

This pattern of results—dictionaries tend to generate broader senses of meaning and corpus linguistics tend to generate narrower senses of meaning, and the sense 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 “a profit arising from office or employ.” But another putative, “broad,” sense need not involve “office” or “employment.”

An impressive analysis of Founding Era dictionaries finds support for the broad interpretation.¹⁶⁸ Across a sample of over one-hundred dictionaries, the broad meaning gains support.

Conversely, a corpus linguistics analysis finds support for a less extensive meaning.¹⁶⁹ 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). 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.

A similar debate characterizes analysis of dictionaries and corpus linguistics concerning 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.¹⁷⁰ 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.”¹⁷¹

These various examples suggest that the processes revealed by the experiments actually manifest in practice. It *is worth noting that we 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 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 corpus linguistics use tends to reflect narrower senses, it could be that in actual legal practice, such initial reflections are

¹⁶⁷ *Carpenter v. United States*, 585 U.S. ___ (2018) (Thomas, dissenting) (citing corpus linguistics data).

¹⁶⁸ John Mikhail, *The Definition of ‘Emolument’ in English Language and Legal Dictionaries, 1523-1806*.

¹⁶⁹ James Cleith Phillips & Sara White, *The Meaning of the Three Emoluments Clauses in the U.S. Constitution: A Corpus Linguistic Analysis of American English, 1760-1799*, 59 S. TEX. L. REV. (2018).

¹⁷⁰ Jack Balkin, *Commerce*, 109 MICH. L. REV. 1 (2010); see also *Gibbons v. Ogden*, 22 U.S. (9 Wheat.) 1 (1824).

¹⁷¹ Randy E. Barnett, *The Original Meaning of the Commerce Clause*, 68 U. CHI. L. REV. 101 (201).

outweighed or overruled—by licit factors including a statute’s precedent or purpose or extrinsic evidence of contracting parties’ intentions, or illicit factors like 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 that 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 seventy or eighty percent of the time), we might expect that citations of dictionary definitions lead to an inclusive interpretation at similar rates (e.g. seventy or eighty percent of the time).

However, one might wonder whether dictionaries are sometimes used politically.¹⁷² If so, we might expect that jurists who cite dictionaries reach narrow/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” and the protection against “cruel and unusual punishments.” Broadly speaking, modern Republicans would prefer the former right interpreted broadly and the latter protection narrowly, while modern Democrats would prefer that the former right is 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 equally in Second and Eighth Amendment cases. First take “keep and bear arms.” The only Supreme Court case in which dictionaries are used to interpret these Second Amendment terms is *District of Columbia v. Heller*.¹⁷³ But this is a rich case. The majority cites dictionaries to interpret all three terms, “keep,” “bear,” and “arms.” And the dissent also cites a dictionary to interpret “bear arms.”¹⁷⁴

First take the majority holding, authored by Republican-appointed Justice Scalia. He cites dictionary definitions of “keep,” “bear,” and “arms.” For “arms,” Scalia cites Samuel Johnson’s 1773 dictionary, which “defined ‘arms’ as ‘[w]eapons of offence, or armour of defense.’”¹⁷⁵ He also cites Timothy Cunningham’s 1771 legal dictionary, which “defined ‘arms’ as ‘any thing that a man wears for his defense, or takes into his hands, or useth in wrath to cast at or strike another.’”¹⁷⁶ Scalia also cited, but does not print, Noah Webster’s 1828 definition.

Scalia also 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.’”¹⁷⁷ Moreover, “Webster defined it as ‘[t]o hold; to retain in one’s power or possession.’”¹⁷⁸

¹⁷² E.g. James J. Brudney & Lawrence Baum, *Dictionaries 2.0: Exploring the Gap Between the Supreme Court and Court of Appeals*, 125 YALE L.J. F. 105 (2015).

¹⁷³ 554 U.S. 570 (2008).

¹⁷⁴ *Id.*

¹⁷⁵ *Id.* at 581.

¹⁷⁶ *Id.*

¹⁷⁷ *Id.* at 285.

¹⁷⁸ *Id.*

Thus, Scalia concludes, “the most natural reading of ‘keep Arms’ in the Second Amendment is to ‘have weapons.’”¹⁷⁹

Finally, Scalia cites Johnson, Webster, Sheridan, and the Oxford English Dictionary for the claim that “bear” meant “carry.”¹⁸⁰ 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, signed by all the Democratically-appointed Justices, cites “bear arms,” finding that its dictionary meaning is “to serve as a soldier, do military service, fight.”¹⁸¹ It also cites the very same Johnson Dictionary definition that Scalia cites—“weapons of offence, or armour of defence”—but understands it to apply narrowly, exclusive of the use contested in *Heller*.

This same pattern emerges in circuit courts.¹⁸² Of all cases citing dictionaries in the same sentence as “keep,” “bear,” or “arms,” one of those defining a Second Amendment term was authored by a Democratic-appointed judge, and six were authored by Republican-appointed judges.

In the one case in which a Democratic-appointed judge used a dictionary, it supported interpreting “bear arms” narrowly.¹⁸³ In the other six, a Republican-appointed judge used a dictionary to interpret keep and bear arms broadly.¹⁸⁴ Thus, at both the Supreme Court and circuit court level, a judge’s dictionary use in Second Amendment cases matched the outcome predicted by the political affiliation of the appointing president.

Contrast this with the use of dictionaries in Eighth Amendment cases. In *Furman v. Georgia*, 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.”¹⁸⁵

However, more recent conservative-authored opinions use dictionaries to construe the Eighth Amendment’s protection narrowly. Consider Thomas and Scalia in *Baze v. Rees*, arguing that lethal injections for executions are constitutional:¹⁸⁶

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”).

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; heard hearted; barbarous. *Dryden*.

¹⁷⁹ *Id.*

¹⁸⁰ *Id.* at 584

¹⁸¹ *Id.* at 646.

¹⁸² I considered all Westlaw-listed cases heard in Federal Courts of Appeal that cite the Second Amendment. I searched within those for uses of “dictionary” within the same sentence as “keep,” “bear,” or “arms.” Seven cases were returned, six of which used a dictionary to define a Second Amendment term.

¹⁸³ *Silveira v. Lockyer*, 328 F.3d 567 (2003).

¹⁸⁴ *Kolbe v. Hogan*, 813 F.3d 160 (2016); *Parker v. District of Columbia*, 478 F.3d 370 (2007); *Nordyke v. Kind*, 319 F.3d 1185 (2003); *Silveira v. Lockyer*, 312 F.3d 1052 (2002).

¹⁸⁵ 408 U.S. 238, 312 (1972).

¹⁸⁶ 553 U.S. 35, 97 (2008).

2. [Of things.] Bloody; mischievous; destructive. *Psalms*.

Now consider “cruel” in Webster’s 1828:

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

It is striking that Scalia and Thomas use the definitions that are applied to persons, rather than the definition 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” punishment 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 (i.e. exclusively/narrowly) when defining “unusual.” For example, in *Harmelin*, the court considered whether the imposition of mandatory sentences of life in prison without the possibility of parole, without any consideration of mitigating factors, constituted cruel and unusual punishment. Scalia writes for the majority, concluding that such punishment is not “unusual.” Unusual means, according to Scalia “such as [does not] occur in ordinary practice, Webster’s American Dictionary (1828), “[s]uch as is [not] in common use,” Webster’s Second International Dictionary 2807 (1954).¹⁸⁷

Finally, in *Farmer* and *Helling*,¹⁸⁸ the Republican-appointed opinions indicate a narrow dictionary construal of “punishment.” It does not include an attack on a prisoner, and punishment is only the penalty for the commission of a crime, not jail conditions.

This same pattern is consistent with the limited evidence from circuit courts.¹⁸⁹ 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*, 780 F.2d 645 (1985), in which a Republican-appointed judge cites Johnson’s dictionary to support that “punishment” does not include injuries sustained when a bus to which prisoners were chained caught fire. He argues that the dictionary definition requires that punishment be deliberate or reckless in the criminal law sense, or a “strong sense.”

It is helpful to consider the broader pattern that emerges in these uses of dictionaries to determine whether legal texts are exclusive (narrow construal) or inclusive (broad construal). Recall Figure 11, which suggests that, cross all levels of the judiciary

¹⁸⁷ *Harmelin v. Michigan*, 501 U.S. 957 (1991).

¹⁸⁸ *Farmer v. Brennan*, 511 U.S. 825 (1994); *Helling v. McKinney*, 509 U.S. 25 (1993).

¹⁸⁹ I considered all Westlaw-listed cases heard in Federal Courts of Appeal that cite the Eighth Amendment. I searched within those for uses of “dictionary” within the same sentence as “cruel,” “unusual,” or “punishment.” Seven cases were returned, one of which used a dictionary to define an Eighth Amendment term.

dictionaries tend to admit of “broad” interpretations about 70% of the time and “narrow” interpretations about 30% of the time. *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 caselaw at the Supreme Court and Circuit Court level does not reflect this pattern. Instead, Republican-appointedees 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-appointedees 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.

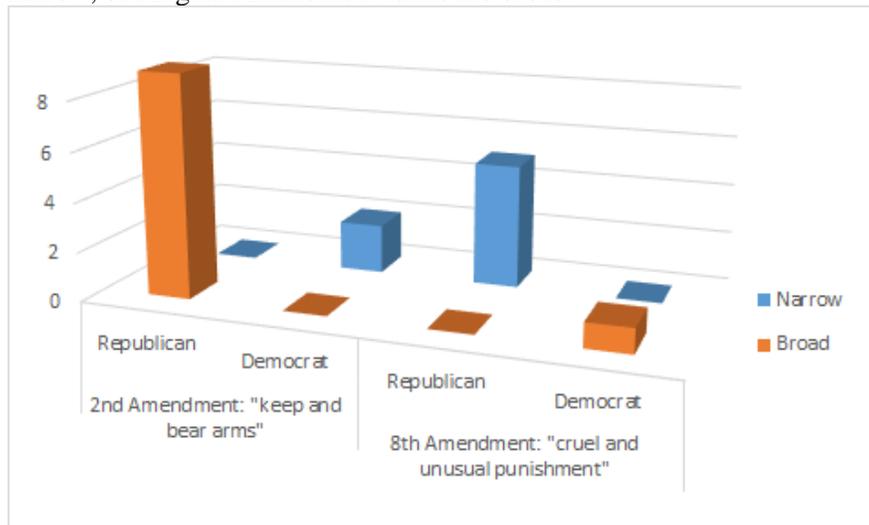


Figure 12. Exclusive (“Narrow”) and Inclusive (“Broad”) Uses of Dictionaries by Republican and Democrat appointed Justices and Judges, in interpreting “keep and bear arms” and “cruel and unusual punishment”

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. Given the novelty of 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. Part VI.A elaborates how the experiments clarify one of the processes underlying reliance on dictionary definitions or word usage data. Specifically, dictionary definitions tend to suggest broad senses of category membership, while word usage data tends to suggest more narrow, prototypical senses of category membership. This is a crucial insight for any theory of legal interpretation that employs these methods as interpretive data; in taking a dictionary definition or pattern of word usage data as interpretive evidence, it is essential to understand what that data commonly tends to suggest to interpreters.

Part VI.B identifies several fallacies of the use of corpus linguistics and dictionaries that are supported by the experimental data. For example, consider “The Non-Appearance Fallacy,” the mistaken assumption that the non-appearance 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 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, ordinary meaning sometimes diverges from ordinary use: People’s full understanding of language is not always reflected in recorded speech and writing, especially their understanding concerning non-prototypical category membership.

Part VI.C considers implications for a set of interpretive theories that rely heavily on dictionary definitions or corpus linguistics to determine legal outcomes. This includes certain formalist, textualist, and originalist views on which a dictionary definition or sets of corpus linguistics data might be taken to be sufficient to determine “the ordinary meaning” of a text and thereby determine the legal outcome. This Part develops a broader burden-shifting argument. The experiments provide evidence that relying solely on dictionaries or 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 non-arbitrary and demonstrably reliable use of corpus linguistics and dictionaries in interpretation.

Finally, Part VI.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 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, the results suggest that dictionary definitions, 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 corpora in interpretation. Verdicts from dictionaries were more strongly correlated with a term’s extensivist uses than its prototypical ones. And verdicts from corpus linguistics were more strongly correlated with a term’s prototypical uses than its extensivist ones.

For example, consider that a car is a prototypical vehicle but an airplane is not (although most people today still judge that an airplane is a vehicle).¹⁹⁰ Most participants using the dictionary were inclined to classify both entities as vehicles, but most participants using the corpus data only classified the (prototypical) car as a vehicle.¹⁹¹

¹⁹⁰ See Section IV.B *supra*.

¹⁹¹ *Id.*

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 (e.g. taking a gun to a gang fight), but more often excluded non-prototypical examples (e.g. driving to a drug deal with a gun in the rear of the car).¹⁹²

These results suggest something about the cognitive mechanisms underlying the use of dictionaries and corpus, respectively. While dictionary definitions help identify more extensive uses of the term, corpus linguistics data tends to help identify prototypical uses. There is something initially puzzling about this finding: Corpus data is 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 (prototypical) 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 broad senses of meaning. Conversely, corpus linguistics as typically practiced in legal contexts—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 non-prototypical ones, such as vehicle-airplanes.

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 two historically famous cases of statutory interpretation problems.¹⁹³ They were chosen for this reason and because they are the first two examples used by the recent manifesto on corpus linguistics.¹⁹⁴ That is, these are examples selected by other researchers who were unaware of the present hypotheses. Moreover, that paper is the leading *defense* of corpus linguistics; insofar as the results here challenge corpus linguistics, I have used cases that should have been selected by its proponents, which should be favorable.¹⁹⁵

But maybe the original paper from which these examples are drawn just happened to select two unusual examples. That remains an open empirical question. The Corpus-Prototypical and Dictionary-Extensivist relationship is certainly somewhat limited, insofar as not all categories have prototypical and non-prototypical members.¹⁹⁶ Thus, the response to this objection is not a flat refutation. Instead, it is an acknowledgement that additional data on other examples may very well enrich our understanding of dictionaries and corpus linguistics further.

Nevertheless, there are some further theoretical 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 corpus linguistics. Advocates of using

¹⁹² *Id.*

¹⁹³ See e.g., Lee & Mouritsen, *supra* note 13; Solan & Gales, *supra* note 45.

¹⁹⁴ Lee & Mouritsen, *supra* note 13.

¹⁹⁵ *Id.*

¹⁹⁶ However, even well-defined terms like “even,” “odd,” “female,” and “plane geometry figure” elicit judgments that feature similar to judgments of prototype concepts. For example, four judged as 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. Sharon Lee Armstrong, Lila R. Gleitman & Henry Gleitman, *What Some Concepts Might Not Be*, 13 COGNITION 265 (1983).

corpus linguistics recommend data from collocation and keywords in context searches. Collocation searches reflect the words that appear most frequently with the relevant search term. 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, prototypical members are the ones with most of all of those features. So, insofar as the statistically common collocates also reflect our core associations with the concept, there is good reason to think that collocation data is especially useful in identifying prototypical category members.

The second type of search, 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 non-prototypical. However, in practice, pragmatic considerations might limit the number of non-prototypical uses that we find.¹⁹⁷

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 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 fact in no way undermines 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 (e.g. 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. 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.

¹⁹⁷ See generally H.P. GRICE, LOGIC AND CONVERSATION (1975).

(4) would be a strange, if not inappropriate, way to describe an airplane. Consider a final example:

(5) How did Jasmin get all those books to school? She carried them there.

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 non-prototypical category members by the broader category description. If you want to point out sharks in the water, you don’t say “look at those big *fish!*”¹⁹⁸ 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 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.

B. Fallacies of Interpretation

This Section identifies fallacies of the use of corpus linguistics and dictionaries, ones made clear by the experimental findings. Of course, there are many other important critiques that are not discussed here.¹⁹⁹ This Section identifies new critiques, grounded in the novel experimental results of this paper.

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

First consider some fallacies of corpus linguistics:

The Non-Appearance Fallacy: The non-appearance 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.²⁰⁰ Defenders of 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

¹⁹⁸ See, e.g., Martin L. Jönsson & James A. Hampton, *On Prototypes as Defaults* COGNITION (2007). But see Andrew C. Connolly, Jerry A. Fodor, Lila R. Gleitman & Henry Gleitman, *Why Stereotypes Don’t Even Make Good Defaults*, 103 COGNITION 1 (2007). 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 non-default interpretations. See generally S.C. LEVINSON, PRESUMPTIVE MEANINGS: THE THEORY OF GENERALIZED CONVERSATIONAL IMPLICATURE (2000).

¹⁹⁹ See, e.g., Solan & Gales, *supra* note 45.

²⁰⁰ See, e.g., *Carpenter v. United States*, 585 U.S. ___ (2018) (Thomas, dissenting) (noting that “[t]he phrase ‘expectation(s) of privacy’ does not appear in . . . the papers of prominent Founders, early congressional documents and debates, collection of early American English texts, or early American newspapers.”)

simpler. . . . we 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*.”²⁰¹

This argument is fallacious. As the experimental results here indicate, corpus linguistics often neglects non-prototypical uses of a term. A corpus search for “vehicle” returns predominantly uses involving cars. But this does not mean that only prototypical returns 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.” This argument is fallacious in our modern moment. It is also fallacious in historical interpretation. A historical corpus is only smaller than the modern corpus used in the experiments, presumably containing even fewer uses (and thus fewer non-prototypical ones).

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

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 corpus linguistics data may not adequately reflect non-prototypical 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.

Consider one final fallacy, the comparative use fallacy.

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 of ordinary meaning.

This fallacy arises when users of 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 this that sense reflects the *only* plain meaning or the “best” meaning.

However, this too is a fallacious argument. Recall the experimental findings. For ordinary people, law students, and United States judges, there were several entities that were classified as vehicles in ordinary language, but not with respect to the corpus data (e.g. airplane, bicycle, electric wheelchair). Imagine there was a debate over the meaning of vehicle. Sense-1 is the inclusive sense (car, truck, airplane, bicycle, and electric wheelchair) and Sense-2 is the exclusive sense (only car and truck; and not airplane, bicycle, or electric wheelchair). Users of corpus linguistics might be inclined to argue that Sense-1 is the better candidate, as it has more support from the corpus. However, this is a fallacious inference. As discussed previously, the omission of non-prototypical uses from the exclusive sense does not mean it is a better sense or one that reflects *the* (only) plain meaning of “vehicle.”

²⁰¹ Lee & Mouritsen, *supra* note 13, at 844.

These three fallacies—The Non-Appearance Fallacy, The Uncommon Use Fallacy, and The Comparative Use Fallacy—each present an individual challenge to common interpretive arguments grounded in corpus linguistics data. But we should also note that these three arguments together threaten much of the usefulness of corpus linguistics. If corpus linguistics cannot reliably exclude omitted or rare uses (from ordinary meaning) or determine which of two possible senses is more credible, this undercuts much of the method’s promise as an independent solution to questions of ordinary meaning.

Now consider two fallacies of dictionary use:

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 meaning. In defining vehicle, we must provide a definition that includes cars and trucks, but also airplanes, submarines, and mopeds. “An agent of transmission or carrier” is helpful in achieving this. But such a broad definition also applies to many entities that are not understood as vehicles. For example, participants predominantly reported that roller skates, baby shoulder-carriers, and zip-lines are *not* vehicles. Yet, many dictionary users categorized these as vehicles. Thus, while it might seem that a dictionary definition is tied tightly to ordinary meaning, this assumption is erroneous. The mere fact that a use “fits” the dictionary definition does not imply that the use is consistent with ordinary understanding. Given the practical nature of dictionaries, aiming to succinctly define 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.

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, but not necessary, criteria of category membership. This is especially common in multi-part 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.²⁰² It is sometimes a mistake to point to a particular aspect of one dictionary definition and argue that any use that does not meet that criteria cannot be part of the ordinary meaning.

²⁰² See Part IV *supra* (ordinary concept participants were divided, about 50-50, that a non-mobile WWII memorial truck was a vehicle; however, dictionary participants overwhelmingly disagreed).

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 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 non-arbitrary methodology.

Some theories of interpretation assume an “ordinary meaning” of terms and phrases, take that as the criterion of legal interpretation, and rely on empirical evidence—for example, a dictionary definition—to determine the meaning and legal outcome. Consider Schwartz and 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.²⁰³

Here the first party 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. Victoria Nourse has documented the increasing tendency for textualist interpreters to rely on the ordinary meaning of specific words in statutes, through textual “gerrymandering,” “intense decontextualization,”²⁰⁴ and “reducing the statute’s meaning to a particular word or two.”²⁰⁵ Moreover, as Abbe Gluck and Lisa Bressman note, these interpretive tasks often concern “very ordinary words.”²⁰⁶ And, in many cases, both “liberal” and “conservative” Justices rely on dictionary definitions to establish the ordinary meaning of these important terms.²⁰⁷

Insofar as the use of dictionary definitions and corpus linguistics is meant to capture how ordinary people would understand particular terms and phrases, the experimental results here indicate that both tools are not accurate 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 corpus linguistics should track the same “ordinary meaning.” The extreme divergences between the use of dictionary definitions and 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

²⁰³ Alan Schwartz & Robert E. Scott, *Contract Interpretation Redux*, 119 YALE L.J. 926, 962 (2010).

²⁰⁴ Nourse, *supra* note 27, at 669.

²⁰⁵ *Id.*, at 681.

²⁰⁶ Gluck & Bressman, *supra* note 46, at 955.

²⁰⁷ See notes 22-26 *supra*.

more from those theories about how, precisely, use of dictionaries and corpus linguistics can accurately achieve that task.

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 these tools perform better in historical analysis, the results provide evidence that these methods are unreliable in historical interpretation:

1. Empirical Claim: The modern uses of dictionaries and 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.
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 method is 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 non-arbitrary and demonstrably reliable methodologies. The two key features of this challenge are non-arbitrariness and demonstrable reliability. I consider those in turn.

First consider non-arbitrariness. There are many choices one must make in interpretation, many of which threaten arbitrariness. 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 corpus linguistics: Exactly how many searches will be conducted, what precisely will be searched and how is the search string determined, what percent of conforming uses “counts” 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 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* corpus linguistics.

The experimental results suggest that corpus linguistics and dictionaries are not just sometimes divergent; they often provide strongly opposing verdicts about ordinary meaning.²⁰⁸ Insofar as a theorist or jurist endorses dictionaries in one instance and corpus linguistics in another—with no further supporting reasons—this raises a new question of arbitrariness. This question becomes more pressing where the choice of methodology seems to match the desired political or legal outcome, such as when judges or interpreters who are quick to point out the absence of (non-prototypical) uses in a corpus in one case, but the breadth of a dictionary definition in another.²⁰⁹

²⁰⁸ See Section IV.E.7 *supra*.

²⁰⁹ Compare, e.g., *Smith v. U.S.*, 508 U.S. 223 241-244 (1993) (Scalia, J., dissenting) (arguing that a broad dictionary definition of “using” should be rejected and instead relying on prototypical examples of “use” in ordinary language to argue that the ordinary meaning of “using” is narrower than the dictionary definition) with *D.C. v. Heller*, 554 U.S. 570 (2008) (Scalia, J.) (arguing that “arms” should be construed extensively on the basis of dictionary definitions).

Thus, there is a burden on theorists who rely upon dictionaries and corpora to elaborate and defend a non-arbitrary use of their tools. Reporting dictionary definitions and detailed corpus data often conveys 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.

This first burden is relatively easier to satisfy. 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, is more demanding. If dictionary definitions and corpus linguistics methods are unreliable, it does not matter much that they are applied systematically. We can construct many non-arbitrary methods of interpretation (e.g. principled dice-rolling is not arbitrary). But any such method is unconvincing until it is also shown 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 corpus linguistics. Overall, the 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 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, Felix 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.”²¹⁰ 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.”²¹¹

For theories and theorists that are 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 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,²¹² 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 corpus linguistics sometimes track *one* of these criteria—often dictionaries track the extensive criterion and corpus linguistics tracks the prototypical one—but the question remains: Which of these—if either—*should* serve as a criterion in legal interpretation?

²¹⁰ Felix Frankfurter, *Some Reflections on the Reading of Statutes*, 47 COLUM. L. REV. 527, 528 (1947).

²¹¹ *Id.*

²¹² See sources cited, notes 91-93, *supra*.

We might consider these two criteria—a broad criterion and a “prototypical” one—against common values that motivate ordinary meaning analysis. Most of those values do not point strongly in favor of either 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 its prototypical sense. Is it more likely that using an extensive criterion in statutory interpretation will lead to more robustly “public” laws?²¹³ 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.”²¹⁴

The previous Section discussed certain formalist, textualist, and originalist theories that often operate as if dictionaries and 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 corpus linguistics, and ordinary understanding—those views have the 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 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”²¹⁵: 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. Corpus linguistics data about prototypical uses could serve as useful evidence confirming inclusion under *ejusdem generis*. For example, finding that “vehicle” refers to busses would be evidence in favor of interpreting a statute concerning “cars, trucks, and other vehicles” to include busses.

In some other circumstances, use of both dictionaries and corpus linguistics would be instructive. For example, consider criminal contexts in which the rule of lenity applies.²¹⁶ Insofar as dictionaries and corpora provide evidence about different senses of a term in this context (e.g. 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

²¹³ E.g. Antonin Scalia, *The Rule of Law as a Law of Rules*, 56 U. CHI. L. REV. 1775 (1989) (arguing that citizens should be able to read statutes and understand what law requires of them).

²¹⁴ ANTONIN SCALIA, *A MATTER OF INTERPRETATION: FEDERAL COURTS AND THE LAW* 23 (1997).

²¹⁵ LLC v. Greater N.Y. Mut. Ins. Co., 31 A.D.3d 100, 103; see also, e.g., Gooch v. United States, 297 U.S. 124, 128 (1936); Aspen Advisors v. United Artists Theater Co., 861 A.2d 1251 (Del. 2004).

²¹⁶ I.e. there is a textual ambiguity and neither of the two possible senses is inconsistent with legislative intent).

definition or corpus linguistics dataset concerning a single term or phrase would not be sufficient to determine the interpretive outcome.

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 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 support that a use is within (at least a non-prototypical) sense of the ordinary meaning.

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 widely register any means of transportation, the broad criterion indicated by a dictionary would be more instructive.²¹⁷ Conversely, if a contract’s purpose is to provide limited insurance for the use of common “vehicles,” the prototypical criterion indicated by the corpus linguistics data may be more instructive.²¹⁸

These considerations about canons, context, and purpose indicate that dictionary definitions and 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 law. Legal interpreters will likely have to look beyond the dictionary and corpus—to the legal text’s context, history, and purpose; and to their other interpretive commitments.

CONCLUSION

This Article has developed a novel “experimental jurisprudence” method of testing two of the fundamental tools of ordinary meaning analysis—dictionary definitions and patterns of word usage through corpus linguistics.²¹⁹ A series of experiments examined judgments of ordinary people, “elite-university” law students, and United States judges,

²¹⁷ ESKRIDGE JR., *supra* note 15, at 23 (“text and purpose are like the two blades of a scissors; neither does the job without the operation of the other”); ESKRIDGE, JR., GLUCK & NOURSE, *supra* note 62, at 6 (2014) (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).

²¹⁸ RESTATEMENT (SECOND) OF CONTRACTS § 202(1) (“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”); W.W.W. Assocs., Inc. v. Giancontieri, 77 N.Y.2d 157, 162 (1990) (a court reads an agreement “as a whole to determine its purpose and intent”).

²¹⁹ For other recent examples of experimental jurisprudence, see Ivar R. Hannikainen & Raff Donelson, *Fuller and the Folk: The Inner Morality of Law Revisited*, in 3 OXFORD STUDIES IN EXPERIMENTAL PHILOSOPHY (forthcoming) (on the ordinary concept of law); James Macleod, *Ordinary Causation*, 94 IND. L.J. (2019) (on the ordinary concept of causation); Christian Mott, *Statutes of Limitations and Personal Identity*, in 2 OXFORD STUDIES IN EXPERIMENTAL PHILOSOPHY 243 (2018) (on the ordinary concept of identity); Roseanna Sommers, *Commonsense Consent*, (draft manuscript) (on the ordinary concept of consent); Joshua Knobe & Scott Shapiro, *What Cognitive Science can Teach Us about Proximate Causation* (draft manuscript) (on the ordinary concept of causation); Kevin P. Tobia, *How People Judge What is Reasonable*, 70 ALA. L. REV. 2915 (2018) (on the ordinary concept of reasonableness); and Kevin P. Tobia, *Legal Concepts and Legal Expertise* (draft manuscript) (on the ordinary and expert-legal concepts of intentional action).

providing evidence bearing on the process and reliability of dictionary and corpus use in interpretation.

The results reveal several common fallacies of interpretation. As one example, recall “The Non-Appearance Fallacy,” the mistaken assumption that the non-appearance of some use in a corpus indicates that this use is outside of ordinary meaning. However, as the experimental results indicate, ordinary meaning sometimes diverges from ordinary use: People’s full understanding of language is not always reflected in recorded speech and writing, especially their understanding concerning non-prototypical category membership.

For certain formalist, 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 corpus linguistics carry serious risks of diverging from ordinary understanding. These results shift the argumentative burden to theorists and practitioners that rely on these tools to determine legal outcomes: In light of the data, these views must articulate and demonstrate a non-arbitrary 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 more narrow, prototypical criterion. Although dictionaries and 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, 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.