very least, it suggests that standard single-corpus analysis is unreliable and prone to cherry-picking, and that researchers should always analyze more than one corpus at a time. More broadly, it implies a kind of meta-indeterminacy, suggesting that even words with a determinate interpretation in one context or in the experience of a specific judge may have different meanings in other contexts or according to other judges.

This Article makes several contributions to the existing literature, both theoretical and empirical. It develops the theory of textual determinacy and uses it to motivate the application of computational methods to augment our understanding of legal text. Using a newly created database of word embeddings, the Article then develops novel computational techniques to understand word meaning and quantify clarity. The Article applies these methods to uncover widespread textual indeterminacy in real-world cases as well as substantial and unexplored variation between corpora. These findings underscore the importance of non-textual evidence in legal interpretation, contrary to interpreters who rely on text alone.

I. A THEORY OF TEXTUAL CLARITY

How do judges currently apply the concept of textual clarity? This Part describes theories of textual clarity and attempts to empirically study the concept under the status quo. It advances a new theory of clarity that explicitly separates knowledge and determinacy, arguing that the focus of many theorists on interpreters' knowledge causes them to overlook insurmountable linguistic indeterminacy.

A. The Use Theory, Textual Clarity, and Determinacy

The basic philosophy underlying empirical analysis of legal text is the "use theory" of meaning, which holds that the meaning of a word is determined by its use. On this view, words are solely distinguished by the contexts in which they can be appropriately applied. If any instance of "automobile" can be replaced with "car" (and vice versa), then their meanings are identical. Similarly, words like "car" and "iguana" differ not because they represent different underlying platonic concepts, but because you wouldn't drive an "iguana" and you wouldn't own a pet "car."[25]

Intuitively, the use theory corresponds with our everyday experiences. A reader might be completely thrown the first time she encounters an unfamiliar word. But after the second, third, or fourth time she sees it, she begins to form a mental model of that word's use; after seeing that word many times, she has a complete picture of its meaning. This also explains why dictionaries, although purporting to present objective definitions of word meaning, still explain those definitions and justify them with reference to examples of actual usage.

In statutory interpretation today, the use theory has completely supplanted the old "representational theory" of interpretation,[26] which "presupposed that the statutory text could have an intrinsic meaning that Congress simply enacted into law. . . . Instead, practically everyone now accepts the insight that language has meaning only because it reflects practices and conventions shared by a community of speakers and listeners."[27]

The use theory powerfully explains our intuitive understandings of textual meaning. But how important in real-world cases is word meaning alone? Legal interpretation is complex, and even the strictest modern textualists don't simply analyze the meanings of isolated words. When appropriate, both textualists and purposivists will consult other interpretive aids, including legislative history, the context of a statute's enactment, and interpretive developments after the initial enactment.[28] They primarily disagree over when these tools should be used, not whether the tools should be used at all.[29]

In choosing between textual and nontextual sources, the most common approach is that courts will follow the text if it's "plain" or "clear," but will incorporate additional evidence (legislative his-

[25] *See generally* Anat Biletzki & Anat Matar, *Ludwig Wittgenstein*, STANFORD ENCYCLOPEDIA OF PHIL. (last updated Oct. 20, 2021), https://perma.cc/VW5L-FGQC (contrasting the use theory with representational theories).

[26] *See infra* Part IV.C (discussing an example where the use and representational theories of meaning diverge).

[27] MANNING & STEPHENSON, *supra* note 3, at 184.

[28] HENRY M. HART, JR. & ALBERT M. SACKS, THE LEGAL PROCESS 1379–80 (William N. Eskridge, Jr. & Philip P. Frickey eds., 1995).

[29] John F. Manning, *What Divides Textualists from Purposivists?*, 106 COLUM. L. REV. 70, 78, 92–93 (2006).

tory in statutory interpretation, extrinsic evidence in the interpretation of contracts, wills, and deeds) otherwise.[30] Of course, the plain meaning rule in turn relies on the determination of whether text is clear—and judges are notoriously oblique about their standards for textual clarity.[31]

In recent years, Justice Brett Kavanaugh has emerged as one of the most prominent critics of clarity doctrines.[32] Justice Kavanaugh has argued that judges disagree on the level of clarity required to declare text clear,[33] and that even if they did agree, determining textual clarity "is often not possible in any rational way."[34] Because textual clarity is a fuzzy concept, Justice Kavanaugh and others have worried that judges will often make clarity determinations on other, less appropriate, grounds.[35] As Professor Ward Farnsworth, J.D. Candidate Dustin Guzior, and Professor Anup Malani have argued, "judgments about ambiguity . . . are dangerous, because they are easily biased by strong policy preferences that the makers of the judgments hold."[36]

[30] *See supra* notes 2–7 and accompanying text. Legal text might be unclear either unintentionally, as a natural byproduct of the drafting process, or strategically, perhaps in order to facilitate compromise. *See generally* Gillian K. Hadfield, *Weighing the Value of Vagueness: An Economic Perspective on Precision in the Law*, 82 CALIF. L. REV. 541 (1994) (discussing the potential for strategic ambiguity in legal drafting); Jeffrey K. Staton & Georg Vanberg, *The Value of Vagueness: Delegation, Defiance, and Judicial Opinions*, 52 AM. J. POL. SCI. 504 (2008) (discussing strategic vagueness in judicial opinions).

[31] *See, e.g.*, Lawrence M. Solan, *Pernicious Ambiguity in Contracts and Statutes*, 79 CHI.-KENT L. REV. 859, 866 (2004) ("[D]ifferent approaches to ambiguity . . . would simply not survive if we were not generally uncertain about what we mean when we talk about ambiguity.").

[32] *See, e.g.*, Kavanaugh, *Fixing Statutory Interpretation*, *supra* note 10, at 2138–39 (criticizing judicial reliance on clarity doctrines as subjective and ambiguous); Brett M. Kavanaugh, *Keynote Address: Two Challenges for the Judge as Umpire: Statutory Ambiguity and Constitutional Exceptions*, 92 NOTRE DAME L. REV. 1907, 1912 (2017) [hereinafter Kavanaugh, *Keynote Address*] ("[T]here is no real objective guide for determining whether a statute is ambiguous.").

[33] Kavanaugh, *Fixing Statutory Interpretation*, *supra* note 10, at 2137 ("One judge's clarity is another judge's ambiguity. It is difficult for judges (or anyone else) to perform that kind of task in a neutral, impartial, and predictable fashion.").

[34] *Id.*

[35] *See, e.g.*, *id.* at 2138–39 ("Because judgments about clarity versus ambiguity turn on little more than a judge's instincts, it is harder for judges to ensure that they are separating their policy views from what the law requires of them.").

[36] Ward Farnsworth, Dustin F. Guzior & Anup Malani, *Ambiguity About Ambiguity: An Empirical Inquiry into Legal Interpretation*, 2 J. LEGAL ANALYSIS 257, 290 (2010); Kavanaugh, *Fixing Statutory Interpretation*, *supra* note 10, at 2138 ("For making that determination, no theory helps; it is simply a judgment about the clarity of the English and whether it is reasonable to read it more than one way. . . . [T]he theories themselves are incapable of generating answers." (quoting Farnsworth et al., *supra*, at 274)).

This Article evaluates both of Justice Kavanaugh's concerns. It uses statistical methods to produce a quantitative, "rational" method for determining clarity in legal text. Then, by empirically applying that method to real-world cases, it considers whether judges assess textual clarity consistently, thereby studying whether legal cases are amenable to judgments of textual clarity at all.

As a matter of theory, we should first separate two distinct aspects of clarity: information and determinacy. Both contribute to judicial findings of textual clarity, meaning the circumstances in which text alone decides the outcome of a case.[37] There are two reasons why a judge might find legal text unclear. First, although she might believe a particular interpretation to be the best, she might lack the *information* necessary to be confident in her judgment. This is a matter of "epistemic limitations as opposed to metaphysical indeterminacy"[38]—perhaps because the judge is cautious about the limits of her own reasoning, or perhaps because she lacks access to tools (e.g., dictionaries, research databases) that might clarify textual meaning.

A second reason that text might not be clear is *indeterminacy*. While inadequate information is a property of the particular judge, indeterminacy is a property of the *text itself*.[39] If linguistic meaning is indeterminate, then all the information and research in the world couldn't shed light on the correct interpretation of a word. In applying legal tests of clarity, two readings may be so close to equally plausible that there would be no point in declaring one of them clearly correct.

[37] This follows the theories of "modified textualists" like Professor Abbe Gluck, who consider statutory text first and consider legislative history only if the text is ambiguous. *See generally* Abbe R. Gluck, *The States as Laboratories of Statutory Interpretation: Methodological Consensus and the New Modified Textualism*, 119 YALE L.J. 1750 (2010). *See also* Kavanaugh, *Fixing Statutory Interpretation*, *supra* note 10, at 2118 ("Several substantive principles of interpretation—such as constitutional avoidance, use of legislative history, and *Chevron*—depend on an initial determination of whether a text is clear or ambiguous."); Lawrence B. Solum, *The Interpretation-Construction Distinction*, 27 CONST. COMMENT. 95, 97 (2010). *See generally* Re, *supra* note 1.

[38] Re, *supra* note 1, at 1511 n.42.

[39] *See generally* Solum, *supra* note 37 (describing "underdeterminacy," where a text is underdeterminate if it admits of more than one possible meaning); Randy E. Barnett, *Interpretation and Construction*, 34 HARV. J.L. PUB. POL'Y 65, 68 (2011) (same). Professor Solum also described a "construction zone" (analogous to the zone of indeterminacy in this Article) in which the text is underdeterminate and "construction (that goes beyond direct translation of semantic content into legal content) is required." Solum, *supra* note 1, at 108.

Indeterminacy can have different sources, the most prominent being ambiguity and vagueness. Text is ambiguous if it could potentially have more than one meaning—for example, the phrase "light baseball cap" is ambiguous as to whether the cap is light in weight or light in color. Text is vague if its limits are imprecisely defined—for example, the Sahara Desert is clearly "hot" and Antarctica is clearly "cold," but there is no clear dividing line between hot and cold, and San Jose might be considered either depending on context. In these cases, no interpreter could resolve the ambiguity or vagueness, regardless of the amount of data she brought to bear.

To be concrete, consider the classic example: Hart's famous "vehicles in the park" hypothetical.[40] A local park contains a sign saying "No vehicles in the park." The sign clearly prohibits cars, and equally clearly allows pedestrians. But what about, say, bicycles? Bicycles seem like vehicles in some respects (they have wheels; they carry people) and unlike vehicles in other respects (they're smaller than cars and are human-powered).[41]

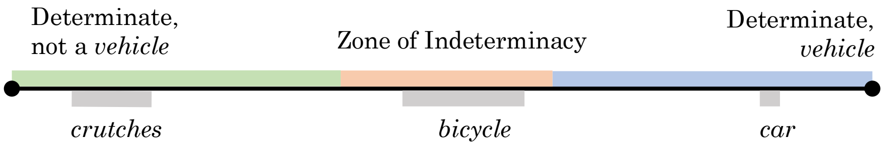
Now imagine that a judge has some internal threshold for textual determinacy,[42] and she will declare text clear if sufficiently confident that the text crosses this threshold. The two variables here are the *threshold* she uses (relating to determinacy), and her *confidence* level (relating to information). Figure 1 depicts this graphically, using Hart's vehicles-in-the-park hypothetical.

[40] *See* Hart, *Positivism*, *supra* note 18, at 607.

[41] Hart described legal language as "open-textured" when it possesses a "fringe of vagueness" that makes its application to borderline cases indeterminate. H.L.A. HART, THE CONCEPT OF LAW 120–25 (1961). *See generally* Brian Bix, *H.L.A. Hart and the "Open Texture" of Language*, 10 L. & PHIL. 51 (1991) (discussing Hart's concept of open texture and its origins in the work of philosophers Friedrich Waismann and Ludwig Wittgenstein).

[42] External thresholds are also possible. *See, e.g.*, Farnsworth et al., *supra* note 36, at 289 (describing external judgments of clarity in terms of predictability). A similar model could apply to external thresholds; there the analog to determinacy would be the average opinion among some group (perhaps laypeople if ordinary meaning is in question) about some textual judgment, and the analog to information would be variance among laypeople in opinions about meaning.

FIGURE 1: HYPOTHETICAL SCALE OF TEXTUAL DETERMINACY



The black line is a continuum representing the degree to which some *x* is a "vehicle." The green and blue shaded areas above the line represent zones in which the judge believes that text alone should be decisive (i.e., the text is clear), assuming complete information.[43] The orange shaded area above the line, which I call the "zone of indeterminacy," is the area where the judge would find text unclear, even given perfect information. The shaded grey areas below the line represent the judge's hypothetical judgments about how strongly crutches, a bicycle, or a car are "vehicles." These areas are confidence intervals rather than points, because the judge isn't completely confident in her judgment. Thus Figure 1 incorporates both indeterminacy (reflected by the orange area) and incomplete information (reflected by the grey areas).

The most important feature of Figure 1 is the orange area, the zone of indeterminacy. It reflects the legal judgment that sometimes cases are too close to be decided on the basis of text alone—not simply as a matter of information or measurement, but because language can be fundamentally indeterminate.[44] Because the confidence interval for "bicycle" overlaps with the zone of indeterminacy, the judge would declare that case unclear and then proceed to apply other evidence—like the legislative history of the sign, the context of enactment, or pragmatic considerations. On the other hand, because the intervals for "crutches" and "car" fall entirely outside the zone of indeterminacy, the judge would declare those cases clear on the basis of text alone.

[43] Professors Richard Re and Ryan Doerfler have argued that clarity thresholds should vary depending on the circumstances of the legal test. Re, *supra* note 1, at 1519; Ryan D. Doerfler, *How Clear Is Clear?*, 109 VA. L. REV. 651, 673–75 (2023).

[44] Lawrence M. Solan, *Learning Our Limits: The Decline of Textualism in Statutory Cases*, 1997 WIS. L. REV. 235, 243–62 (describing the limits of language in deciding legal cases).

The zone of indeterminacy helps to explain many ongoing debates in legal interpretation. One way to characterize the distinction between textualists and purposivists in statutory and constitutional interpretation is that textualists have a narrower zone of indeterminacy—they're willing to accept word meaning as decisive even when a purposivist might declare the text ambiguous and use legislative history to break the tie.[45] Similarly, a way to distinguish formalists from contextualists in contract interpretation and patent litigation is that formalists have a narrower zone of indeterminacy as well, making them less willing to consult extrinsic evidence.[46] The cutoff between determinacy and indeterminacy will vary from interpreter to interpreter; the evidence that an interpreter will use when ambiguity exists will also vary.

To be sure, textual analysis may be useful even when indeterminate—textual meaning could just be one factor weighed alongside other interpretive considerations. But text is clearly *less* useful in those cases. So it may be unsurprising that textualists often try to minimize the appearance of textual indeterminacy. Arch-textualist Justice Antonin Scalia, for example, has argued that the bicycle-vehicle comparison is determinate—that a bicycle is in fact not a vehicle.[47]

More generally, Justice Scalia has said that textual meaning "usually . . . is easy to discern and simple to apply,"[48] and Justice Neil Gorsuch has agreed that "[s]tatutory ambiguities are less like dandelions on an unmowed lawn than they are like manufacturing defects in a modern automobile: they happen, but they are pretty rare."[49] The stakes of this view are high. Believing that text is usually clear, textualists are more likely to reject deference to

[45] For example, in *Exxon Mobil Corp. v. Allapattah Services, Inc.*, 545 U.S. 546 (2005), noted purposivist Justice John Paul Stevens argued that the relevant statute was ambiguous, *id.* at 575 (Stevens, J., dissenting), while Justice Anthony Kennedy, whose interpretive views leaned more toward textualism, asserted that the statute in question was not ambiguous, *id.* at 567 (majority opinion). Another way to characterize the difference is that textualists are systematically more confident in their personal interpretations of text or have better information about text; however, it's not clear why this would be so.

[46] Alan Schwartz & Robert E. Scott, *Contract Interpretation Redux*, 119 YALE L.J. 926, 962 (2010); Peter Lee, *Patent Law and the Two Cultures*, 120 YALE L.J. 2, 30 (2010).

[47] ANTONIN SCALIA & BRYAN A. GARNER, READING LAW: THE INTERPRETATION OF LEGAL TEXTS 37–38 (2012).

[48] ANTONIN SCALIA, A MATTER OF INTERPRETATION: FEDERAL COURTS AND THE LAW 45 (1997).

[49] NEIL GORSUCH, A REPUBLIC, IF YOU CAN KEEP IT 136 (2019) (quoting Judge Raymond Kethledge).

agencies under the *Chevron* doctrine,[50] to uphold patents against claims of indefiniteness,[51] to adopt literal interpretations of statutes,[52] and more. But because little empirical scholarship has considered the nature of textual clarity, these claims have largely gone unchallenged so far.

B. Empirical Analysis of Legal Text in the Status Quo

Judges have long recognized the value of consulting outside sources in textual interpretation, rather than relying solely on personal intuition. Modern courts often rely on dictionaries to elucidate unclear language, as the Supreme Court has done regularly since the mid-1800s.[53] But commentators have frequently criticized dictionary use: scholars argue that judges can "dictionary shop" for the definition that best suits their preferred outcome, because dictionaries contain so many competing definitions.[54] Dictionary editors themselves have condemned the use of dictionaries by courts—the editor at large of the *Oxford English Dictionary* has said that "it's probably wrong, in almost all situations, to use a dictionary in the courtroom."[55] Moreover, because dictionaries provide only broad guidance on the meanings of words, judges must still weigh competing definitions both within

[50] In *Chevron* step one, a court engages in ordinary statutory interpretation to determine if the statute is clear. Chevron U.S.A., Inc. v. Nat. Res. Def. Council, 467 U.S. 837, 842–43 (1984). Although textual clarity is just one aspect of the inquiry, if the text is clear then typically the statute will be considered clear as a whole.

[51] *See* 35 U.S.C. § 112.

[52] A vast literature criticizes literal interpretations of statutory text in different areas of law. *See, e.g.*, Lawrence Zelenak, *Thinking About Nonliteral Interpretations of the Internal Revenue Code*, 64 N.C. L. REV. 623, 637 (1986) (criticizing textualism in tax law); Daniel J. Bussel, *Textualism's Failures: A Study of Overruled Bankruptcy Decisions*, 53 VAND. L. REV. 887, 896–99 (2000) (criticizing textualism in bankruptcy law); Bradford C. Mank, *Is a Textualist Approach to Statutory Interpretation Pro-Environmentalist?: Why Pragmatic Agency Decisionmaking Is Better Than Judicial Literalism*, 53 WASH. & LEE L. REV. 1231, 1241 (1996) (criticizing textualism in environmental law).

[53] Note, *Looking It Up: Dictionaries and Statutory Interpretation*, 107 HARV. L. REV. 1437, 1454 (1994). The Supreme Court first mentioned a dictionary in 1785, in *Respublica v. Steele*, 2 U.S. (2 Dall.) 92, 92 (1785) (discussing a litigant's citation of author Samuel Johnson's *Dictionary of the English Language*, which was originally published in 1755). *Id.* at 1437 n.2.

[54] *See generally* Ellen P. Aprill, *The Law of the Word: Dictionary Shopping in the Supreme Court*, 30 ARIZ. ST. L.J. 275 (1998). *See also* James J. Brudney & Lawrence Baum, *Oasis or Mirage: The Supreme Court's Thirst for Dictionaries in the Rehnquist and Roberts Eras*, 55 WM. & MARY L. REV. 483, 566 (2013) (suggesting that Justices may selectively report one of several definitions offered within a single dictionary in order to justify their decision).

[55] Adam Liptak, *Justices Turning More Frequently to Dictionary, and Not Just for Big Words*, N.Y. TIMES (June 13, 2011), https://perma.cc/6YKQ-QQ9E.

each dictionary and between different dictionaries to arrive at a highly subjective and opaque judgment of textual clarity.

Reacting to criticism of dictionaries, scholars and judges have recently begun to use corpus linguistics to provide quantitative evidence of textual meaning. Corpus linguists consult databases of real-world language use to draw conclusions about how words are used in real life. For example, in *Smith v. United States*,[56] the Supreme Court considered a statute that imposed a thirty-year mandatory minimum sentence on any defendant who "during and in relation to any crime of violence or drug trafficking crime . . . uses . . . a firearm."[57] Should the mandatory minimum apply to a defendant who traded a firearm for drugs? Professors Stefan Gries and Brian Slocum, two legal corpus linguists, answered this question by searching a corpus for instances of the word "use" to see how often it denoted a trade. They found that applicable instances of "use" never involved a trade or barter (at least in the corpus they chose); this evidence suggested that the mandatory minimum should not have applied.[58]

Corpus linguistics suffers from a wide variety of problems, which scholars have commented on elsewhere.[59] One problem that scholars have not extensively explored is that because corpus linguistics focuses on simple word frequencies, it misses important aspects of semantic meaning. For example, imagine that a statute addresses the "driver of any train, aircraft, automobile, or other mode of transportation." Is a jet pilot a "driver" under this statute? Many corpus linguists would answer the question by searching a corpus for the words around "pilot" and "driver."[60] "Pilot" might co-occur with words like "aircraft," "airport," and "tarmac"; "driver" might co-occur with words like "automobile," "garage," and "road." Because these co-occurring words have little overlap

[56] 508 U.S. 223 (1993).

[57] 18 U.S.C. § 924(c)(1) (1992).

[58] Stefan Th. Gries & Brian G. Slocum, *Ordinary Meaning and Corpus Linguistics*, 2017 B.Y.U. L. REV. 1417, 1461–62.

[59] *See, e.g.*, Tobia, *supra* note 11, at 757 (describing how traditional corpus linguistics is generally "underinclusive" with respect to legal questions). *See generally* Anya Bernstein, *Legal Corpus Linguistics and the Half-Empirical Attitude*, 106 CORNELL L. REV. 1397 (2021) (criticizing traditional corpus linguists for selectively adopting an empirical approach to the law); Lawrence M. Solan, *Can Corpus Linguistics Help Make Originalism Scientific?*, 126 YALE L.J. F. 57 (2016) (acknowledging the promise of legal corpus linguistics while also noting potential limitations).

[60] This is known as the "collocation" method. Thomas R. Lee & Stephen C. Mouritsen, *Judging Ordinary Meaning*, 127 YALE L.J. 788, 831–32 (2018).

on a simple frequency analysis, the traditional corpus linguist might conclude that a pilot is not a type of "driver."

However, these simple frequencies don't adequately encode *semantic meaning*—the meaning that captures the essential relationship between words. "Aircraft" and "automobile," "airport" and "garage," and "tarmac" and "road" are close semantic analogs, differing in their superficial context (planes versus cars) rather than their underlying meaning. These contextual differences don't demonstrate that a pilot is not a kind of "driver"; yet these contextual differences are exactly where corpus linguistics directs our attention.

The semantic blind spot inherent in corpus linguistics isn't a niche problem limited to contrived hypotheticals—it affects virtually every corpus linguistics analysis, sometimes dramatically. A focus on frequencies will tend to depress estimates of similarity, in turn leading to excessive false negatives. In *Smith*, "uses" is a broad term that could have been intended to capture "trades" (as the Supreme Court held it did[61]), just as "driver" is a broad term that could have been intended to capture "pilot." By focusing on simple analysis of word frequencies, corpus linguistics ignores this nuance.

Another problem with corpus linguistics is that it demands a wide variety of judgment calls behind its veneer of scientism and objectivity,[62] which can make its results seem more determinate than they really are. Applying the theoretical framework from the previous Section, corpus linguists frame textual clarity as primarily a question of measurement. On that view, uncertainty exists because of incomplete information, and corpus data can "help us resolve different types of linguistic uncertainty in the interpretation of legal texts."[63] Figure 2 illustrates this perspective, showing

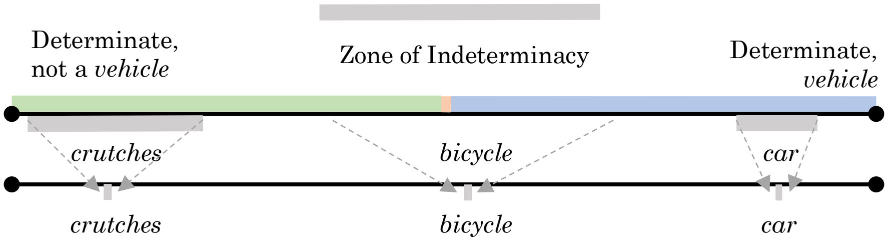
[61] *Smith*, 508 U.S. at 225.

[62] For descriptions of corpus linguistics as "scientific," see, for example, Clark D. Cunningham & Jesse Egbert, *Scientific Methods for Analyzing Original Meaning: Corpus Linguistics and the Emoluments Clauses* 5–8 (Ga. St. U. College of Law, Legal Studies Research Paper No. 2019-02, 2019) (available at https://ssrn.com/abstract=3321438); Brief of Professors Clark D. Cunningham & Professor Jesse Egbert as Amici Curiae in Support of Neither Party at 5–11, *In re* Trump, 958 F.3d 274 (4th Cir. 2020) (No. 18–2486); Carissa Byrne Hessick, *More on Corpus Linguistics and the Criminal Law*, PRAWFSBLAWG (Sept. 11, 2017), https://prawfsblawg.blogs.com/prawfsblawg/2017/09/more-on-corpus-linguistics-and-the-criminal-law.html (featuring an extended back-and-forth between corpus linguist Stephen Mouritsen and corpus linguistics–skeptic Carissa Hessick over whether corpus linguistics is "scientific").

[63] Lee & Mouritsen, *supra* note 60, at 829; *see also id.* at 851 ("Do corpus data yield means of measuring ordinary meaning? We think the answer is a resounding yes."). Some

how the confidence interval for each word theoretically narrows after the introduction of corpus evidence. This hypothetical corpus linguist has a narrow zone of indeterminacy (consistent with the textualist leanings of most corpus linguists) but wider confidence intervals for each word prior to the application of corpus data. Before applying corpus data, the interpreter would find it unclear whether a bicycle is a vehicle or not; after applying corpus data, the interpreter would find that a bicycle is clearly a vehicle.

FIGURE 2: A CORPUS LINGUIST'S HYPOTHETICAL SCALE OF TEXTUAL DETERMINACY, BEFORE AND AFTER CORPUS DATA



This approach creates several problems. An emphasis on measurement overlooks basic indeterminacy, even though indeterminacy is the way that linguists usually think about ambiguity and vagueness.[64] Ambiguity and vagueness aren't just quirks of inadequate datasets; they're fundamental features of our language. A phrase like "light baseball cap" is simply textually indeterminate—there's no way to adjudicate whether the hat is light in color or weight based on language alone.

The practical impact of a narrow zone of indeterminacy is to make results highly sensitive to subjective methodological choices. It's easy to nudge the confidence interval for "bicycle" across the zone of indeterminacy if that zone is small. Because it's rarely obvious which choice is best as a matter of theory, the results of corpus linguistics hinge on decisions that seem trivial and

scholars have adopted this view too, although they're the exception rather than the rule. *See* Ryan D. Doerfler, *High-Stakes Interpretation*, 116 MICH. L. REV. 523, 527 (2018) ("[T]o say that the meaning of a statute is 'clear' or 'plain' is, in effect, to say that one *knows* what the statute means." (emphasis in original) (citation omitted)).

[64] Confusingly, "most of the discussion in the case law . . . treats the terms 'ambiguous' and 'vague' as synonymous (denoting lack of clarity)." MANNING & STEPHENSON, *supra* note 3, at 274.

whose importance the interpreter may not realize, or worse, may consciously exploit.[65]

Outside of corpus linguistics, quantitative empirical studies have so far provided little guidance on the question of textual clarity. While scholars have surveyed ordinary people to evaluate the meanings of statutes[66] and contracts,[67] these studies have generally not considered the possibility that these documents might be unclear. They instead assume sharp cutoffs in textual interpretation—for example, a bicycle is a "vehicle" if some set percentage of respondents agree that it is, and is not a "vehicle" otherwise.[68] The one study to explicitly assess textual clarity and ambiguity so far did so as an inquiry into bias in judicial decision-making, finding that perceptions of clarity were significantly biased by interpreters' policy preferences.[69] But by focusing on judicial decision-making, this study made a similar move to corpus linguists—it considered textual clarity as an epistemic matter rather than a feature of the text itself.

Overall, then, existing empirical work leaves significant room for improvement in our understanding of legal text. Dictionaries are imprecise and problematic; corpus linguistics is highly subjective, fails to account for relevant semantic meanings, and has so far focused excessively on questions of information rather than indeterminacy. Quantitative empirical studies either assume away the question of textual clarity entirely, or else (like corpus linguistics) focus on issues of information and bias rather than the possibility of indeterminacy.

[65] In statistical terms, the problem is that corpus linguists employ "syntactic context," "semantic context," and "pragmatic context" to limit their searches in ways that reduce the sample size (*N*) of their ultimate inquiry. *See* Lee & Mouritsen, *supra* note 60, at 821–24 (describing how to apply these types of context to "limit our search"). Small-*N* searches are more prone to give widely varying results based on small changes in parameters.

[66] Tobia, *supra* note 11, at 773–77.

[67] Ben-Shahar & Strahilevitz, *supra* note 11, at 1782.

[68] Tobia, *supra* note 11, at 773–74; Ben-Shahar & Strahilevitz, *supra* note 11, at 1779–80.

[69] Farnsworth et al., *supra* note 36, at 271 (surveying law students on whether they believed that text from legal cases was clear or unclear). The study also found these biases could be mitigated by encouraging respondents to take an "external" perspective, where they imagined the views of others rather than relying on their own views. *Id.* at 276; *see also* Lawrence Solan, Terri Rosenblatt & Daniel Osherson, Essay, *False Consensus Bias in Contract Interpretation*, 108 COLUM. L. REV. 1268, 1292–95 (2008) (finding that judges and laypeople disagreed about the correct interpretation of contractual text in an experimental setting, and taking this as indirect evidence of ambiguity).

II. EMPIRICAL METHODS

Although corpus linguistics has become increasingly influential in real-world courts,[70] scholars continue to criticize it as arbitrary and unreliable.[71] Corpus linguists have responded that even though their techniques are imperfect, the interpretation of text is a core judicial task, and critics have offered no better alternative. "It takes a method to beat a method,"[72] they argue, and criticism in the absence of proposed solutions is unconstructive.

Part I of this Article stated the problem; the remainder proposes a solution. It offers a new computational method, providing two main actionable improvements over status quo approaches. First, the Article quantifies the degree of semantic indeterminacy in legal text as a general matter, suggesting that most cases are semantically indeterminate and that text alone doesn't provide a clear answer. The natural solution is for judges to rely less on legal text and more on other extrinsic evidence, like legislative history or tiebreaker rules like substantive canons of construction. Second, the Article proposes tools that could be used to assess textual determinacy and investigate textual meaning in individual cases. While these tools warrant further study and can't yet be used in all cases,[73] they represent an important first step toward improving or replacing corpus linguistic methods.

A. Word Embeddings

Over the past decade, artificial intelligence researchers have made huge advances in the field of natural language processing, which uses computational models to analyze language. One of the

[70] *See, e.g.*, Murray v. BEJ Mins., LLC, 464 P.3d 80 (Mont. 2020); Health Freedom Def. Fund v. Biden, 599 F. Supp. 3d 1144, 1160 (M.D. Fla. 2022) (using corpus linguistics analysis as evidence for determining the meaning of "sanitation" when considering the CDC's authority to impose a mask mandate).

[71] Their foremost complaint is that corpus linguistics focuses only on "prototypical" meanings and therefore produces excessive false negatives, a phenomenon sometimes described as the "Nonappearance Fallacy" in which people erroneously reason that because a meaning doesn't appear in a corpus it isn't legitimate. Tobia, *supra* note 11, at 734–35. Corpus linguists have responded that their methods are more flexible than critics have appreciated and can address this critique. Thomas R. Lee & Stephen C. Mouritsen, *The Corpus and the Critics*, 88 U. CHI. L. REV. 275, 331–32 (2021). While this is true, flexibility brings its own problems, as Part II describes.

[72] Lee & Mouritsen, *supra* note 71, at 351.

[73] *See infra* Part IV.D.

most significant advances has been the development of embedding models.[74] These models have revolutionized natural language processing. They're responsible for a wide range of recent innovations, including language models like ChatGPT[75] and surprising improvements to translation tools like Google Translate.[76]

Broadly speaking, an embedding model represents words as mathematical vectors, known as "word embeddings." Each vector will have many dimensions, typically hundreds, with each dimension intuitively reflecting one aspect of a word's semantic meaning. An embedding model begins with an optimization problem, attempting to produce the vectors for each word that best explain the empirical distribution of the words in a real-world corpus.[77] In the process of optimization, the model compresses co-occurrence statistics into a multidimensional representation of each word that captures core aspects of semantic meaning. Thus, the word embeddings represented by the vectors give a richer sense of meaning than simple word frequencies.

These word embeddings encode semantic distinctions in useful and intuitive ways that corpus linguistics can't account for. The vector space generated by an embedding model includes predictable geometric relationships between related pairs—between

[74] *See* Tomas Mikolov, Kai Chen, Greg Corrado & Jeffrey Dean, Efficient Estimation of Word Representations in Vector Space 2–4 (Sept. 7, 2013) (conference paper) (available at https://perma.cc/RY8F-HVDC) (introducing model architectures for computing word embeddings).

[75] *Introducing ChatGPT*, *supra* note 13.

[76] Gideon Lewis-Kraus, *The Great A.I. Awakening: How Google Used Artificial Intelligence to Transform Google Translate, One of Its Most Popular Services—And How Machine Learning Is Poised to Reinvent Computing Itself*, N.Y. TIMES MAG. (Dec. 14, 2016), https://perma.cc/KU6A-SSXT.

[77] For example, Google's popular Word2vec model is essentially a neural network that takes as an input any word in the vocabulary, and outputs a probability distribution corresponding to the likelihood that other words in the vocabulary co-occur with the query word within a given context window (say, three words on either side). *See generally* Mikolov et al., *supra* note 74 (introducing the Word2vec model). Stanford University's Global Vectors for Word Representation (GloVe) model attempts to optimize an objective function based on the likelihood that any two words will co-occur. The GloVe model used throughout this Article was designed in order to generate geometric relationships between words that facilitate analogical reasoning. *See generally* Pennington et al., *supra* note 12. GloVe models are marginally more popular among social scientists. Moreover, GloVe underweights rare terms while Word2vec underweights common ones, which means that "Word2Vec is likely to be less 'robust,' that is, embeddings will tend to be more corpus specific, than GloVe." Pedro L. Rodriguez & Arthur Spirling, *Word Embeddings: What Works, What Doesn't, and How to Tell the Difference for Applied Research*, 84 J. POL. 101, 111 (2022).

zip codes and cities, companies and their CEOs, and more.[78] Word embeddings can also capture analogistic relationships between different words—for example, by showing that:

$$\overrightarrow{Par\imath s} - \overrightarrow{France} + \overrightarrow{Japan} \approx \overrightarrow{Tokyo}$$

(The arrow above each word indicates that it's a vector; for example, $\overrightarrow{Par\imath s}$ is the vector for the word "Paris.")

As Part II.D illustrates, this sort of vector algebra provides reassurance that word embeddings are encoding meaningful textual relationships. It also helps to explain the differences between words, which usefully complements otherwise opaque similarity metrics.

B. Cosine Similarity

Is a judge a "representative[ ]" whose election is governed by federal law?[79] Are fossils "minerals" the ownership of which transfers with oil and gas rights?[80] Is a tomato a "vegetable[ ]" subject to a higher tariff rate?[81] Legal cases frequently turn on whether some x is a y. These are essentially questions of semantic similarity, a classic task for word embedding models.[82] Graphically, we can see this in the angles between different vectors generated by a word embedding model, where a smaller angle implies that the vectors are more similar. Figure 3 shows one hypothetical vector space, compressed from the hundreds of dimensions typically used in word embeddings to two dimensions.

[78] Jeffrey Pennington, Richard Socher & Christopher D. Manning, *GloVe: Global Vectors for Word Representation*, STANFORD UNIV. (Aug. 2014), https://perma.cc/7V95-DDM6.

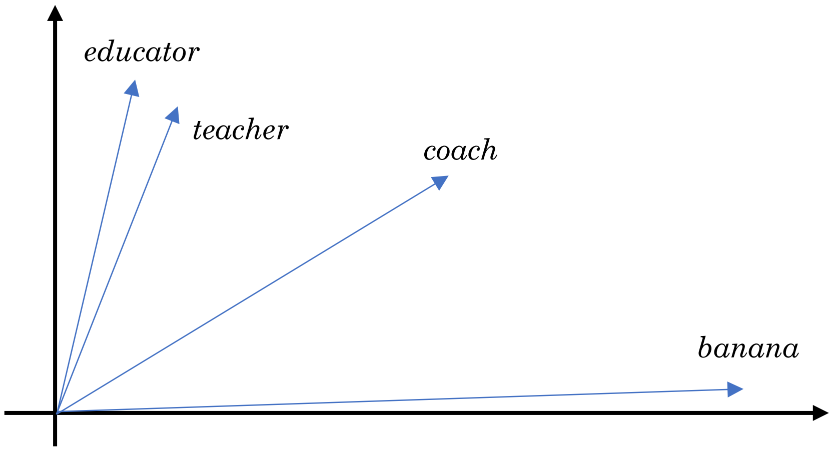
[79] Chisom v. Roemer, 501 U.S. 380, 389 (1991).

[80] Murray v. BEJ Mins., LLC, 924 F.3d 1070, 1072 (9th Cir. 2019) (en banc).

[81] Nix v. Hedden, 149 U.S. 304, 306 (1893).

[82] These are sometimes also described as "hyponym/hypernym" questions, asking whether some term (the hyponym) fits within the broader category represented by some other term (the hypernym). Rion Snow, Daniel Jurafsky & Andrew Y. Ng, *Learning Syntactic Patterns for Automatic Hypernym Discovery*, 2004 PROC. 17TH INT'L CONF. ON NEURAL INFO. PROCESSING SYS. 1297, 1297–98, 1300–02. For example, if a fossil is a mineral, we could think of "fossil" as the hyponym and "mineral" as the hypernym.

FIGURE 3: HYPOTHETICAL WORD EMBEDDING MODEL



In Figure 3, the angle between $\overrightarrow{\textit{teacher}}$ and $\overrightarrow{\textit{educator}}$ is very small, reflecting the fact that they're synonyms. The angle between $\overrightarrow{\textit{teacher}}$ and $\overrightarrow{\textit{coach}}$ is slightly wider, indicating that they're similar but perhaps not synonyms, and the angle between $\overrightarrow{\textit{teacher}}$ and $\overrightarrow{\textit{banana}}$ is wider still, indicating that they have very different meanings. In the literature on word embeddings, this angle is quantified as the "cosine similarity" between two vectors. A cosine similarity of 1, the maximum possible, denotes identical vectors, while a cosine similarity of -1, the minimum, denotes exact opposites.[83] In practice, most cosine similarity estimates fall between 0 and 1.[84]

Because word embeddings are constructed by evaluating words' near neighbors, a cosine similarity using word embeddings intuitively reflects the degree to which two words could be substituted for each other in ordinary text. "Educator" could sensibly substitute in almost all situations where "teacher" appears; "coach" is an adequate but imperfect substitute; and "banana" is a nonsensical substitute. As we'll see, this intuitive explanation

[83] Mathematically, vectors that are completely orthogonal will have a cosine similarity of zero.

[84] Because the cosine similarity intuitively measures how plausible word *x* would be as a replacement for word *y* in any given sentence, antonyms generally have a higher cosine similarity than words that are totally unrelated. For example, in the English Wikipedia, "good" and "evil" have an average cosine similarity of 0.306, while "good" and "cucumber" have an average cosine similarity of 0.099.

for the cosine similarity closely matches prevailing theories of legal textual interpretation.[85]

C. Benchmarking Cosine Similarity

Word embeddings and cosine similarity help to formalize textual interpretation, but they aren't particularly useful on their own. A judge who heard that the cosine similarity between two words is 0.35 would have no idea whether the words are very similar, not similar at all, or somewhere in between. It takes additional work to translate quantitative cosine similarities into *qualitative* legal judgments.

One simple method might be to assign a cutoff, so that, for example, words would be deemed similar if their cosine similarity exceeded 0.5 and dissimilar otherwise. Past studies have used this approach to interpret statutes[86] and contracts[87] by surveying ordinary people for their intuitions on word meaning.[88] (In those

[85] While word embeddings and cosine similarity are well suited to hyponym-hypernym inquiries, we should exercise caution in extending them to word similarity in other domains. The cosine similarity between, for example, "Subaru" and "Volvo" in the English Wikipedia illustrates the perils of relying on cosine similarity in other kinds of investigations: for those words, it's 0.360224 (averaged over bootstraps), higher than the cosine similarity for "wheelchair" and "vehicle," even though a Subaru is clearly not a Volvo. This is because "Subaru" could sensibly substitute in many (but not all) sentences where "Volvo" is used. (Sensibly in "I drove my Volvo to the store," but insensibly in "Volvo is a brand from Sweden.") This seems to be the wrong result.

The issue relates to the study of linguistic taxonomy, an approach to computational linguistics that preceded recent advances in artificial intelligence. *See* FRIEDRICH UNGERER & HANS-JÖRG SCHMID, AN INTRODUCTION TO COGNITIVE LINGUISTICS 64–66 (2d ed. 2006); Michael Gasser, *How Language Works: Word Senses and Taxonomies*, IND. UNIV. (2022), https://perma.cc/M7NU-N7QL. In a linguistic taxonomical tree, "Volvo" and "Subaru" both are children of "car" (because they're both *types* of cars), which in turn is a child of "vehicle" (because a car is a type of vehicle), etc.

The problem is that cosine similarity only produces meaningful results to questions involving two words on the same branch of a taxonomical tree. Thus, a cosine similarity can sensibly answer whether a Subaru is a "car" or a Subaru is a "vehicle," but not whether a Subaru is a Volvo. Because the ultimate question answered by cosine similarity is whether *x* could substitute for *y*, there are situations where *x* could substitute for *y* even if they're clearly dissimilar—as described in note 84, antonyms generally have a higher cosine similarity than words that are totally unrelated. This Article considers only "is-a" questions where both words are from the same taxonomical branch, and it recommends applying the methods in this Article only to those cases. Fortunately, real-world cases are generally of this sort. Indeed, all of the real-world cases discussed in this Article are arguably examples of direct hyponym-hypernym comparisons.

[86] Tobia, *supra* note 11, at 773–77.

[87] Ben-Shahar & Strahilevitz, *supra* note 11, at 1779–80.

[88] The only study to my knowledge that attempts to explicitly quantify clarity or lack thereof in legal language is Farnsworth et al., *supra* note 36. Farnsworth et al. surveyed almost one thousand law students on whether they believed that text from legal cases was

studies, an *x* is considered a *y* if and only if some set percentage of respondents agree that it is.)

The cutoff method has two important weaknesses. First, it assumes that all text has a determinate meaning (i.e., that there's no zone of indeterminacy). If a land deed entitles a litigant to "minerals,"[89] for example, then the cutoff approach will determine that her right to own fossils on that land turns solely on whether "fossil" is sufficiently similar to "mineral." Second, the cutoff approach is arbitrary. Readers might disagree about whether the correct cutoff to classify a bicycle as a "vehicle" is 30% of cases in agreement, or 40%, 50%, 60%, or 70%.

Of course, not even the most radical textualists rely on word meaning to the exclusion of all other considerations. Any realistic empirical method should include the zone of indeterminacy and should also account for different zones of indeterminacy between different interpreters. This is a better model for how judges and scholars actually think, but a more difficult one to theorize. Compounding the difficulty, cosine similarities have no set meaning: a cosine similarity of 0.5 might be on the high end for one corpus but only in the middle for another. The empirical range of cosine similarities from training a word-embedding vector space will depend on the hyperparameters of the training methodology and many other factors.

Rather than choosing an arbitrary cutoff, this Article introduces a new method to convert quantitative cosine similarities to qualitative assessments of similarity: benchmarking against an established similarity scale. In particular, this Article uses H.L.A. Hart's famous "vehicles in the park" hypothetical by taking a list of potential synonyms for "vehicle" and ranking them according to their cosine similarity. The words include some intuitively similar to "vehicle," like "car" and "automobile," but also some that are intuitively dissimilar, like "skates" and "crutches," with many words in between. Then, rather than simply reporting the cosine similarity between any given *x* and *y*, this Article situates that cosine similarity on the vehicle scale, establishing that *x* and *y*

clear or unclear. *Id.* at 257. However, this study primarily analyzed the presence of bias in judgments about clarity (finding significant bias based on policy preferences) as well as the benefit of encouraging an "internal" perspective based on personal judgments, versus an "external perspective" where a respondent would imagine the likely views of others (finding that the external perspective nearly eliminated bias). *Id.* at 257–59. Thus Farnsworth et al. studied judicial behavior and decision-making bias rather than textual ambiguity itself.

[89] *Murray*, 924 F.3d at 1072 (describing the typical language in a mineral deed).

are as similar as, say, "vehicle" and "bicycle." For example, it might conclude that "fossil" and "mineral" are approximately as similar as "vehicle" and "bicycle."

The vehicle scale serves several purposes. Most fundamentally, it provides an intuitive interpretation of unintuitive quantitative results. Another alternative to the vehicle scale would be testing for statistical significance, for example by testing whether the cosine similarity between "fossil" and "mineral" is greater than between an average random pair of words.[90] But this alternative has limited practical value. Assuming that "fossil" and "mineral" are more similar than a random pair of words (which is virtually inevitable), this alternative doesn't tell us whether they're as similar as, say, "skis" and "vehicle," or as similar as "car" and "vehicle."[91] In contrast, the vehicle scale helps to establish not just the existence of similarity but also the *degree* of similarity.

Moreover, the vehicle scale helps to validate the computational methodology. An interpreter can look at the scale itself to see whether the ordinal ranking of similarities corresponds with her own intuitions. The vehicle scale discussed below in Part III.A is reassuringly sensible, ranking "car" as more vehicle-like than "bicycle," which in turn is more vehicle-like than "skis."

Finally, the scale allows the interpreter to quantify her own zone of indeterminacy in order to compare it against novel word pairs. If the interpreter feels that "bicycle" is neither decisively similar nor decisively dissimilar to "vehicle," then another word pair with the same level of cosine similarity as bicycle-vehicle will also be presumptively indeterminate. This can help individual interpreters to explore the consistency of their intuitions between different contexts.

D. Advantages of Computational Methods

The computational methods described above have several advantages over the corpus linguistics techniques that dominate the status quo. First, computational methods produce a richer account of semantic meaning than corpus linguistics. Unlike corpus

[90] Stefan Th. Gries, *Corpus Linguistics and the Law: Extending the Field from a Statistical Perspective*, 86 BROOK. L. REV. 321, 347–48 (2021) (applying this approach).

[91] The threshold for two words to be more similar than *two randomly selected words* is very small, given that most words have no semantic relationship. Gries's research suggests that in general the cosine similarity between two randomly selected words is zero. *Id.* at 348.

linguistics analysis, embedding models capture the semantic similarity between terms like "driver" and "pilot," despite scant overlap in their co-occurring words.

Second, computational methods permit fewer degrees of freedom than corpus linguistics. Corpus linguists will conduct an initial search of the corpus, but will slice down that search based on their (sometimes contradictory) judgments of which search results are most relevant.[92] Even then, they will sometimes analyze only some subset of the possible search results if there are too many to feasibly read. After identifying the relevant subset, corpus linguists must decide on which of several methods to use and must individually read lines within the corpus to decide how to categorize each line. As Part III.B illustrates with a case study, each step in this process is fraught with subjective judgment calls.

Computational methods reduce opportunities for subjective judgments, as compared to both corpus linguistics and informal textualist methods. Corpus linguists may freely switch between corpora and methods; informal textualists cherry-pick even more aggressively by quoting only the subset of the relevant corpus or dictionary that suits their preferred interpretation.[93] In contrast, this Article uses a single method (cosine similarity) and can apply it to several corpora at once, including by quantifying differences between corpora, as discussed in Part III.E. This explicitly addresses the degrees of freedom that are possible by shopping between different corpora, an important source of variation that can be quantified using computational methods.

Third, computational methods allow us to quantify word similarity, especially in indeterminate cases. As noted above, corpus linguists and textualists have both generally played down the possibility of textual indeterminacy. But it's difficult to know whether they've done so because they believe that clarity is truly pervasive as an objective empirical matter or because they have idiosyncratic thresholds for textual clarity. That is, do textualists believe that most legal cases are as clear as asking whether a car is a "vehicle," and that other ordinary English speakers would agree? Or do they believe that most cases are like asking whether

[92] *See* Carissa Byrne Hessick, *Corpus Linguistics and the Criminal Law*, 2017 BYU L. REV. 1503, 1522 ("At the very least, the different approaches that Gries and Slocum take to determine the most frequent meaning of 'harbor' and the most frequent meaning of 'use' demonstrate that humans must make these choices and that true linguistic experts will sometimes take different approaches to limiting their search results.").

[93] *See* Cary Franklin, *Living Textualism*, 2020 SUP. CT. REV. 119, 125–26 (discussing this style of "[l]iving [t]extualism").

a bicycle is a "vehicle," and that even if most readers would find this comparison unclear, textualists believe the answer is clear? Computational methods, and particularly the vehicle scale, allow us to separate the two.

E. Explaining and Editing Word Vectors

Critics often describe modern machine learning techniques as opaque, and word embeddings are no exception. In contrast, cosine similarity compresses many dimensions of semantic meaning into a single, easily understood number. But what if some dimensions matter more than others? Cosine similarity tells us how different two words are, but not *why* they're different.

Enter vector algebra. Word embeddings naturally lend themselves to analogistic reasoning, which can be quantified as vector algebra.[94] Consider again the analogy between national capitals:

$$\overrightarrow{Paris} - \overrightarrow{France} + \overrightarrow{Japan} \approx \overrightarrow{Tokyo}$$

We can confirm this formula by actually calculating the vector for $\overrightarrow{Paris} - \overrightarrow{France} + \overrightarrow{Japan}$ through vector algebra, and then seeing which word vector has the highest cosine similarity with the resulting vector. As expected, the answer is $\overrightarrow{Tokyo}$, with cosine similarity of 0.833.

In addition, vector algebra is sufficiently flexible that it can identify the words closest to the difference between two single word vectors. While this method is less precise, it can illuminate the difference between two-word vectors in general terms.[95] For example, we can ask:

$$\overrightarrow{fossil} - \overrightarrow{mineral} \approx ?$$

[94] Tomas Mikolov, Wen-tau Yih & Geoffrey Zweig, *Linguistic Regularities in Continuous Space Word Representations*, 2013 Proc. NAACL-HLT 746, 746–47. *See generally* Carl Allen, Ivana Balaževic & Timothy Hospedales, *What the Vec?: Towards Probabilistically Grounded Embeddings*, 32 Proc. 33rd Int'l Conf. on Neural Info. Processing Sys. 7465 (2019).

[95] Professors Alex Gittens, Dimitris Achlioptas, and Michael Mahoney discussed vector additivity and proved theoretically that it should be expected to hold under a set of assumptions that plausibly applies in common word embedding models. *See generally* Alex Gittens, Dimitris Achlioptas & Michael W. Mahoney, *Skip-Gram – Zipf + Uniform = Vector Additivity*, 1 Proc. 55th Ann. Meeting Ass'n Computational Linguistics 69 (2017).

TABLE 1: NEAREST NEIGHBORS FOR $\overrightarrow{fossil} - \overrightarrow{mineral}$

Word	**Cosine Similarity**
mosasaur	0.537
crocodilians	0.520
sauropod	0.498
pterosaurs	0.494

These results suggest that, within this corpus, the primary difference between a fossil and a mineral is that a fossil is associated with dinosaurs. We can conduct the same exercise with any word pair, illuminating words' semantic differences. Section 0 of the Appendix includes nearest neighbors from vector subtraction for all of the real-world cases discussed in this Article.

Simple vector algebra provides background for computational results, and it provides some reassurance that those results reflect real differences in semantic meaning. More intriguingly, it suggests a method to edit word vectors to pinpoint the most legally relevant dimensions of meaning. Section 0 of the Appendix proposes a new method to refine word vectors in this way.

While methods to edit word vectors could be used to focus on particular aspects of meaning in legal text, these methods are relatively new and would benefit from additional research. I describe them only as an initial first step in considering how the computational tools in this Article could apply to subtler legal questions. Editing vectors adds another source of subjective variation, undermining one key advantage of the computational approach. Consequently, these methods should be used sparingly, and their use raises important philosophical questions about contextualism in general.

F. Textualism and Contextualism

A key question underlying much of the discussion so far is how context should inform the meaning of text. For simplicity, most of the example analyses in this Article are acontextual. That is, instead of using context to narrow the meaning of a statutory term like "firearm," I initially analyze the unmodified meaning of "firearm," and then use contextual evidence if and only if that analysis delivers unclear results.

Judges often disagree on whether to apply "formalist" interpretation that focuses on the meaning of specific words or "contextualist" interpretation that takes broader semantic, syntactic, or pragmatic context into account. In several recent high-profile Supreme Court cases, majorities and dissents have both used textual analysis to reach opposite legal conclusions, crucially depending on which contextual information they decided to include.[96] On one hand, there is often good linguistic reason to incorporate context, which can help illuminate the communicative content of legal text.[97] On the other hand, critics have recently observed that contextualist interpretation gives rise to "textual gerrymandering,"[98] arguing that context makes textualism as malleable as the purposivism whose flexibility new textualists like Justice Scalia had complained about.[99]

A formalist applying the plain meaning rule will first focus on narrow textual meaning, turning to contextual evidence only

[96] *Compare* King v. Burwell, 576 U.S. 473, 486–87 (2015) (taking statutory context into account), *with id.* at 501–02 (Scalia, J., dissenting) (declining to take statutory context into account). *Compare* Bostock v. Clayton County, 140 S. Ct. 1731, 1754 (2020) (declining to take statutory context into account), *with id.* at 1766–67 (Alito, J., dissenting) (taking statutory context into account). Many other recent Supreme Court cases feature similar disagreements between the majority and dissent arising from methodological differences over the correct approach to textual interpretation. *See generally* Brnovich v. Democratic Nat'l Comm., 141 S. Ct. 2321 (2021); Johnson v. Guzman Chavez, 141 S. Ct. 2271 (2021); Borden v. United States, 141 S. Ct. 1817 (2021). *See also* William N. Eskridge, Jr. & Victoria F. Nourse, *Textual Gerrymandering: The Eclipse of Republican Government in an Era of Statutory Populism*, 96 N.Y.U. L. REV. 1718, 1814–25 (2021) (compiling and discussing these cases).

[97] *See, e.g.*, Larry Alexander, *Formalist Textualism and the* Cernauskas *Problem*, 23 J. CONTEMP. LEGAL ISSUES 169, 172 (2021) ("Ascertaining what legislatures are asserting through their texts can never be as algorithmic and free of fallible judgment calls as formalist textualism requires."); Erik Encarnacion, *Text Is Not Law*, 107 IOWA L. REV. 2027, 2056 (2022) (criticizing Justice Gorsuch's formalist *Bostock* opinion as "literalism"). On the other hand, prominent textualist Tara Leigh Grove has resisted the critique that formalist textualism is necessarily narrow or literalistic, arguing that formalists should still take semantic context into account, but not pragmatic context. Tara Leigh Grove, *The Misunderstood History of Textualism*, 117 NW. L. REV. 1033, 1096 (2023).

[98] Eskridge & Nourse, *supra* note 96, at 1721–22 (discussing both "choice of text" and "choice of context" as problems with textual interpretation). This Article explicitly discusses "choice of context" in Part II.F; it also implicitly disallows "choice of text" by focusing on single words, rather than whole phrases. However, choice of text may be a problem for future models that incorporate entire statutory phrases. *See also* Victoria Nourse, *Picking and Choosing Text: Lessons for Statutory Interpretation from the Philosophy of Language*, 69 FLA. L. REV. 1409, 1412, 1409 (2017) (suggesting that the choice to focus on "one piece of text over another can amount to assuming that which one is trying to prove" and "can put the thumb on the scales of any interpretation"); Franklin, *supra* note 93, 126, 136, 141–46, 149–51 (discussing various "shadow decision points" that implicitly affect the outcome of textual analysis but are rarely explicitly discussed).

[99] Eskridge & Nourse, *supra* note 96, at 1724–25.

if the narrow text is unclear.[100] For our purposes, this means a formalist interpreter should analyze text using the narrowest possible block of text—for example, by asking whether dinosaur fossils are "minerals," as opposed to "minerals in, on and under, and that may be produced from the lands."[101] Then, if the narrow language is indeterminate, the formalist interpreter could consider additional evidence of *all* kinds, including textual canons, legislative history, and the interplay between the narrow text and the broader context in which the text was written.[102]

However, the methods in this Article could be applied both by formalist and contextualist interpreters. The prior Section discusses tools that can be used to incorporate context in word vectors or focus on particular aspects of meaning; Part IV.D discusses additional tools that could analyze entire texts rather than specific words, which would accommodate contextualist theories of interpretation. Applying these tools introduces additional subjectivity in textual analysis, but a contextualist might ultimately conclude that the subjectivity is worthwhile to obtain more accurate results.

III. Results and Implications

A. The Vehicle Scale

With this background, we can consider some provisional results from similarity analysis using word embeddings. Table 2 takes the "vehicles in the park" hypothetical, testing the similarity between pairs of words: "vehicle" and "car," "vehicle" and "automobile," etc. The words in the scale were selected because they are common across multiple corpora and provide a smooth gradation of cosine similarities.[103] Section A of the Appendix provides more information about the corpora and methods used.

[100] This framing assumes that the interpreter follows some hierarchical method of interpretation, like Professor John Manning's "new purposivism" or Professor Abbe Gluck's "modified textualism." *See* John F. Manning, *The New Purposivism*, 2011 Sup. Ct. Rev. 113, 129–46; Gluck, *supra* note 37, at 1758.

[101] *Murray*, 924 F.3d at 1072.

[102] Professor Doerfler has suggested that the level of desired clarity should be determined based on "the purposes of the applicable [clarity] doctrine." Doerfler, *supra* note 43, at 658. This approach is also consistent with the methodology described in this Article; I do not take a specific view on how clarity thresholds should be determined, only how to test them once they are determined.

[103] The values are averages generated through bootstrapping—that is, by resampling sentences from the corpus to generate corpora of equivalent size, then retraining the word-

TABLE 2: COSINE SIMILARITY RESULTS FOR THE VEHICLE SCALE

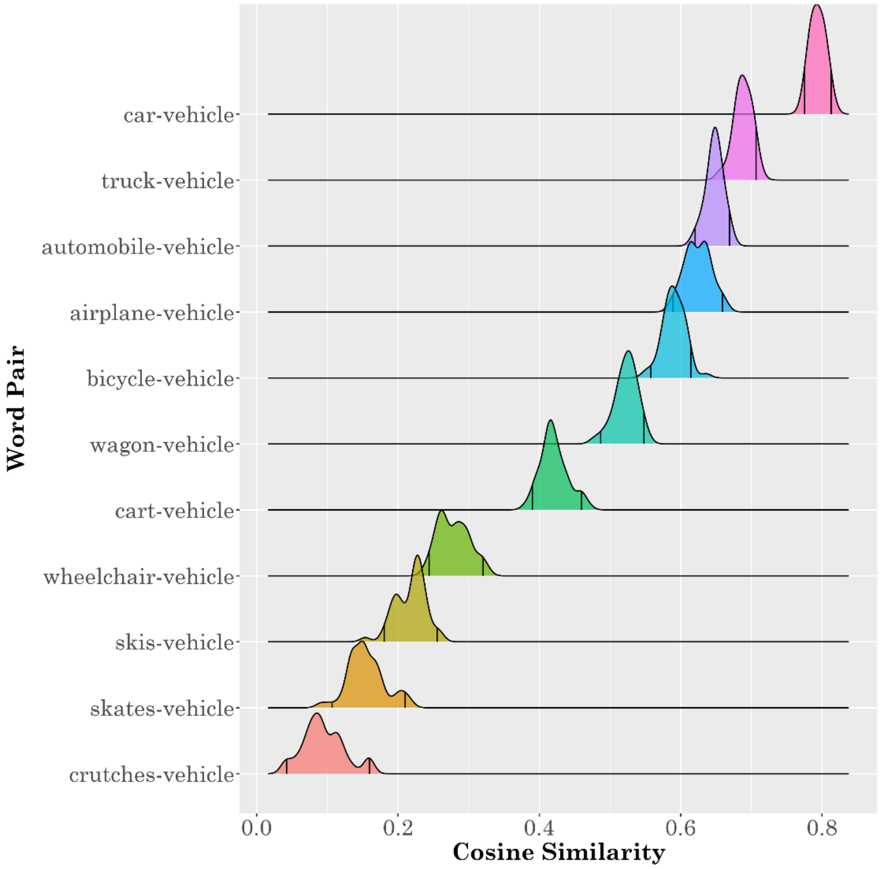
Word	**Cosine Similarity**
car	0.79384
truck	0.688271
automobile	0.647854
airplane	0.624358
bicycle	0.590361
wagon	0.522523
cart	0.420745
wheelchair	0.278497
skis	0.217786
skates	0.155137
crutches	0.094621

The order is intuitive, providing support for the validity of the vehicle scale. In addition to averages, we can also develop a sense of the stability of these estimates and our degree of confidence in the difference between the vehicle candidates by plotting probability distributions for the results.

Figure 4 below displays these plots, including 95% confidence intervals denoted by the black vertical lines inside each curve. It matches the hypothetical scales discussed in Part II, reflecting both determinacy and information. The estimates of cosine similarity represent levels of determinacy, and we could imagine a zone of indeterminacy in the middle of the scale, with the width of the zone varying depending on the preferences of the interpreter. Variation in the estimates of cosine similarity due to incomplete information is reflected in the size of the confidence intervals.

embedding vector using the reconstructed corpora. This was done fifty times using the English Wikipedia, a large corpus with more than four billion words.

FIGURE 4: BOOTSTRAPPED COSINE SIMILARITIES FOR THE VEHICLE SCALE



As discussed above, cosine similarity intuitively reflects how appropriate it would be to replace word *x* with word *y* in a variety of situations. Could the sentence "I parked my *vehicle* at the store" be converted to "I parked my *bicycle* at the store"? What about "I drove my *bicycle* to the store"? Or "She keeps a spare tire in the trunk of her *vehicle*"?

Because "bicycle" appropriately substitutes in some but not all of the sentences where "vehicle" is used, its cosine similarity falls in a middle range. Conversely, "car" can appropriately substitute in almost all sentences where "vehicle" is used, and "crutches" can appropriately substitute in almost none. Their appropriateness as substitutes is reflected in their high and low cosine similarities, respectively, consistent with the use theory of meaning.

B. A Case Study: *Health Freedom Defense Fund, Inc. v. Biden*

Having established the vehicle scale as a benchmark, we can now apply it to real-world cases. Start with a recent, high-profile example: *Health Freedom Defense Fund, Inc. v. Biden*,[104] the infamous[105] Florida district court decision where Judge Kathryn Kimball Mizelle struck down the Biden administration's public transportation mask mandate. One of Judge Mizelle's main arguments concerned the meaning of "sanitation" in the Public Health Services Act of 1944,[106] which empowered the Department of Health and Human Services (HHS), and thus the CDC,[107] to "make and enforce such regulations as . . . are necessary to prevent the introduction, transmission, or spread of communicable diseases."[108] In particular, the Act and its implementing regulations permitted the CDC to provide for "sanitation" measures, which Judge Mizelle identified as the source of the CDC's authority to promulgate a mask mandate.[109]

In Judge Mizelle's view, "sanitation" could be read one of two ways, either reflecting "the sense of cleaning" or "the sense of preserving cleanliness."[110] The former is active, involving "direct cleaning of a dirty or contaminated object," while the latter is passive, involving "a measure to maintain a status of cleanliness, or . . . a barrier to keep something clean."[111] To support her argument, Judge Mizelle analyzed the Corpus of Historical

[104] 599 F. Supp. 3d 1144 (M.D. Fla. 2022).

[105] For legal criticism of Judge Mizelle's textual analysis, see generally Stefan Th. Gries, Michael Kranzlein, Nathan Schneider, Brian Slocum & Kevin Tobia, *Unmasking Textualism: Linguistic Misunderstanding in the Transit Mask Order Case and Beyond*, 122 COLUM. L. REV. F. 192 (2022).

[106] 42 U.S.C. § 264(a).

[107] 42 C.F.R. § 70.2.

[108] 42 U.S.C. § 264(a).

[109] Although I focus on Judge Mizelle's corpus linguistics analysis in this Article, other aspects of her reasoning were even more questionable, including her sidestepping the fact that the statute also permitted the Secretary of HHS to authorize "other measures, as in his judgment may be necessary." 42 C.F.R. § 70.2. This arguably obviated the need to analyze "sanitation" at all, although Judge Mizelle argued that pursuant to the ejusdem generis canon, the "other measures" must be interpreted to be of a character with "sanitation." *Health Freedom Def. Fund*, 599 F. Supp. 3d at 1157–58.

[110] *Id.* at 1159.

[111] *Id.* at 1160.