Brooklyn Law Review

Volume 86 | Issue 2 Article 3

9-22-2021

Big Data and Accuracy in Statutory Interpretation

Brian G. Slocum

Big Data and Accuracy in Statutory Interpretation

Brian G. Slocum[†]

INTRODUCTION

Scholarship is increasingly devoted to improving the "accuracy" of statutory interpretations, but accuracy is a contingent concept dependent on interpretive perspective. If, for instance, a scholar focuses on the language production of the legislature, she may seek to improve the methodology of statutory interpretation through a more sophisticated understanding of the legislative process.[1] Thus, the scholar may argue that one can assess the reliability of the different types of legislative history by focusing on the actors and processes that produce them.[2] Conversely, a scholar might focus on the language comprehension of some speech community,[3] such as the one comprised of "ordinary people."[4] Such a scholar may argue that certain interpretive canons are valid approximations of language usage outside of the law.[5]

Scholars do not normally explicitly identify their work as falling within a certain interpretive perspective, and courts often

[†] Professor of Law, University of the Pacific, McGeorge School of Law. J.D., Harvard Law School; Ph.D. in Linguistics, University of California, Davis. For helpful comments, the author would like to thank the participants at the 'Data-Driven Approaches to Legal Interpretation' symposium at Brooklyn Law School.

[1] *See* Abbe R. Gluck, Comment, *Imperfect Statutes, Imperfect Courts: Understanding Congress's Plan in the Era of Unorthodox Lawmaking*, 129 HARV. L. REV. 62, 62–66 (2015) (analyzing statutory interpretation through the processes used by Congress in enacting legislation).

[2] *See, e.g.*, Jesse M. Cross, *Legislative History in the Modern Congress*, 57 HARV. J. ON LEGIS. 91, 91 (2020).

[3] A "speech community" is a group of people who share a set of linguistic norms and expectations regarding the use of language. *See* Peter L. Patrick, *The Speech Community*, WORKING PAPER, ESSEX RESEARCH REPORTS IN LINGUISTICS (2011).

[4] *See* Bostock v. Clayton Cty., Ga., 140 S. Ct. 1731, 1766 (2020) (Alito, J., dissenting) (arguing that the key question is "[h]ow would the terms of a statute have been understood by ordinary people at the time of enactment?"); *see also* Hillel Y. Levin, *Contemporary Meaning and Expectations in Statutory Interpretation*, 2012 U. ILL. L. REV. 1103, 1103 (arguing that "judges interpreting ambiguous statutes . . . should be constrained by the understanding and expectations of the contemporary public as to the law's meaning and application").

[5] *See, e.g.*, Brian G. Slocum, *Conversational Implicatures and Legal Texts*, 29 RATIO JURIS. 23, 24–25 (2016) (arguing that the *ejusdem generis* canon can be justified based on general linguistic principles).

justify interpretive sources from multiple interpretive perspectives.[6] Nevertheless, if "accuracy" is a coherent concept to apply to a statutory interpretation, some interpretive objective must be identified. The more difficult aspect of accuracy involves measuring a potential statutory interpretation against the interpretive objective. Courts and scholars have not traditionally attempted empirical verification of individual interpretive sources, or the confluence of them in a given interpretation.[7] Thus, appeals to accuracy have been rhetorical rather than empirical.

Consider, for instance, Justice Scalia's appeals to accuracy. In his *Reading Law* book, Justice Scalia argued that "most interpretive questions have a right answer," and thus, "[v]ariability in interpretation is a distemper."[8] Why, then, are judges frequently at odds about which possible interpretation is "correct"? Justice Scalia argued that methodological variation is the cause of interpretive dissensus because not all judges use the "fair-reading method," which gives a text meaning in accordance with "how a reasonable reader, fully competent in the language, would have understood the text at the time it was issued."[9] If judges would use the "fair-reading method," they would "arrive at fairly consistent answers."[10] One problem with Justice Scalia's approach is that the validity of his empirical claims rests on the counterfactual of all judges applying the "fair-reading method."[11] But judges have never all agreed on the same interpretive methodology to interpret statutes, and likely never will.[12] Furthermore, the "fair-reading method," with its non-empirically-based "reasonable reader,"[13] does not foreclose the significant judicial discretion inherent in choosing among available, and

---

[6] For ease of exposition, this Article uses the broad term "interpretive source" to refer to both specific interpretive principles (such as a textual canon) and general sources of information (such as dictionaries or legislative history).

[7] "Empirical verification" of an empirical source involves some attempt to measure its use by the relevant speech community. One notable example involved Abbe Gluck and Lisa Bressman's efforts to survey legislative drafters about whether they consider various interpretive principles when drafting legislation. *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, 904–11 (2013). Empirical verification of a given interpretation involves efforts to consult the relevant speech community regarding the meaning of a particular statute. *See infra* Part IV (discussing the use of surveys to determine how ordinary people interpret statutes).

[8] ANTONIN SCALIA & BRYAN A. GARNER, READING LAW 6 (2012).

[9] *Id.* at 33.

[10] *Id.* at 36.

[11] Of course, the judges would also need to apply the "fair-reading method" correctly.

[12] Richard H. Fallon, Jr., *The Meaning of Legal "Meaning" and Its Implications for Theories of Legal Interpretation*, 82 U. CHI. L. REV. 1235 , 1255–62 (2015) (listing the various ways that "meaning" can be defined).

[13] *See* SCALIA & GARNER, *supra* note 8, at 33.

often conflicting, interpretive sources.[14] Thus, Justice Scalia offers no reason to believe that his approach would uniquely lead to "correct" answers, or even increased interpretive consistency.

In comparison to traditional rhetoric-centric appeals to accuracy, as exemplified by Justice Scalia's above arguments, data-driven approaches to statutory interpretation may reorient how arguments about accuracy are made and evaluated. This essay considers two data-driven approaches to statutory interpretation: surveys and corpus linguistics. The use of surveys is a tool of the growing "experimental jurisprudence" movement that is by definition empirically based.[15] In turn, corpus linguistics typically involves the statistical analysis of data from a corpus, which is a machine-readable "compilation of written and transcribed spoken language used in authentic communicative contexts" (such as in newspapers, novels, books, etc.).[16] If performed competently, corpus linguistics results meet the "scientific standards of generalizability, reliability, and validity."[17] Surveys and corpus linguistics thus have the potential to help judges make more empirically based decisions about statutory meaning, although neither can transform statutory interpretation into an empirical science.

This essay considers data-driven approaches and claims about interpretive accuracy through an evaluation of how these interpretive sources fit within the traditional structures of statutory interpretation. Part I explains that every statutory interpretation is made in light of some objective of interpretation, but these objectives reflect different interpretive perspectives.[18] A basic distinction is between interpretive perspectives that focus on the language production of the legislature and those that focus on the language

---

[14] *See* William N. Eskridge, Jr., *The New Textualism and Normative Canons*, 113 COLUM. L. REV. 531, 531 (2013) (reviewing ANTONIN SCALIA & BRYAN A. GARNER, READING LAW: THE INTERPRETATION OF LEGAL TEXTS (2012)) (pointing out that one of the problems with Justice Scalia's claims is that "[f]or any difficult case, there will be as many as twelve to fifteen relevant 'valid canons' cutting in different directions, leaving considerable room for judicial cherry-picking").

[15] *See, e.g.*, James A. Macleod, *Ordinary Causation: A Study in Experimental Statutory Interpretation*, 94 IND. L.J. 957, 991–1012, 1016 (2019) (analyzing how an ordinary reader would understand Title VII's language by asking ordinary readers to apply that language in context, drawing on a set of nationally representative survey experiments).

[16] *See* Brief for Amici Curiae Corpus-Linguistics Scholars Professors Brian Slocum, Stefan Th. Gries, and Lawrence Solan in Support of Employees at 7, Bostock v. Clayton Cty., Ga., 140 S. Ct. 1731 (2020) (No. 17-1618).

[17] Haoshan Ren et al., *"Questions Involving National Peace and Harmony" or "Injured Plaintiff Litigation"? The Original Meaning of "Cases" in Article III of the Constitution*, 36 GA. ST. U. L. REV. 535, 540 (2020) (emphasis omitted).

[18] Of course, the court will typically not announce the interpretive perspective, and it may shift during the interpretive process.

comprehension of some community of speakers.[19] In turn, as Part II explains, interpretive sources serve a subservient role by providing evidence that helps select the meaning that best satisfies the relevant objective of interpretation. A given interpretive source might provide evidence relevant to one objective of interpretation but not another. For instance, it is intuitive that legislative history provides information relating to the language production of the legislature.[20] Consider, though, that asserting the relevance of evidence from legislative history to the language comprehension of ordinary speakers might require the aid of a legal fiction.[21]

As Part III explains, corpus linguistics, as currently practiced, is an interpretive source relevant to the ordinary meaning objective of interpretation. As such, corpus linguistics provides information about conventions of communication that apply outside of the law, and thus the language comprehension of ordinary people.[22] Contrary to the claims of some proponents of corpus linguistics, though, the integration of corpus analysis into the structure of statutory interpretation does not make the determination of statutory meaning an empirical issue. Unlike various other interpretive sources such as legislative history, the main function of corpus analysis is to provide data about word meanings that cut across contexts.[23] Corpus linguistics can therefore reveal important systematicities of language usage, but any corpus analysis must be combined with an examination of the particularized context of a statute in order to determine the meaning of the relevant provision. Judges must examine this context without the aid of corpus linguistics because replication of the exact language of a statutory provision in a corpus of nonlegal language, let alone the entire context of a statute, is quite unlikely.

Corpus linguistics is thus like other traditional interpretive sources in the sense that it provides indirect evidence of a community's views regarding certain aspects of an interpretive dispute and must be combined with other evidence in

---

[19] *See* William Eskridge, Jr., Brian G. Slocum & Stefan Th. Gries, *The Meaning of Sex: Dynamic Words, Novel Applications, and Original Public Meaning*, 119 MICH. L. REV. 1503, 1516 (2021). Judge Frank Easterbrook, for example, believes that the "significance of an expression depends on how the interpretive community alive at the time of the text's adoption understood those words." *See* SCALIA & GARNER, *supra* note 8, at XXV; *see also* Richard H. Fallon, Jr., *The Statutory Interpretation Muddle*, 114 NW. U. L. REV. 269 285–88 (2019) (explaining how intentionalists and textualists describe the objective of interpretation).

[20] The relevant community of speakers may well include members of the legislature depending on how the community is defined. *See supra* note 7 and accompanying text.

[21] *See infra* notes 60–62 and accompanying text.

[22] *See* Stefan Th. Gries & Brian G. Slocum, *Ordinary Meaning and Corpus Linguistics*, 2017 BYU L. REV. 1417, 1422–33 (2017).

[23] *See infra* note 150 and accompanying text.

order to resolve the dispute.[24] In contrast, as Part IV explains, a survey is a direct way of measuring a particular group's views about an issue and may even be used to measure how ordinary people interpret and apply a specific statute to a particular set of facts. Surveys thus resolve the tension between empiricism and context by being able to measure the views of a particular group with respect to a specific statute and interpretive dispute.[25] Furthermore, when surveys address the ultimate interpretive question, they obviate the judicially determined ordinary meaning standard because there is no need to make inferences from generalizations about language usage.[26]

Even so, surveys likely cannot adequately account for all of the traditional interpretive sources because legal training and knowledge are integral to statutory interpretation. Statutory interpretation is a multilayered process that involves normative decisions, specialized legal competence, and inferences from context. The potential of survey evidence, like corpus linguistics, raises important questions about the empiricism of statutory interpretation but, like in other areas of law, proponents of the empirical view may confuse normative notions with empirical ones.[27] Still, when interpreting statutes, judges often make assertions about an objectified person or community's views about meaning, and survey evidence can help evaluate the accuracy of those assertions.

The continuing adaptation of corpus linguistics and surveys as sources of meaning for legal interpretation is an exciting development. But they must be properly situated within the process and theory of interpretation in order to assess whether they can transform how the accuracy of statutory interpretations is measured. Both corpus linguistics and surveys have the potential to help judges make more empirically based decisions about statutory meaning. Even so, features of legal interpretation prevent either source from transforming legal interpretation into an empirical science. Legal knowledge and training, the full context of a statute, and interpretive inferences and judgment are all integral to statutory interpretation, and these features prevent statutory interpretation from being fully informed by empirical methods. Appeals to interpretive accuracy must therefore remain at least partly rhetorical.

---

[24] *See infra* notes 126–150 and accompanying text.
[25] *See infra* notes 158–160 and accompanying text.
[26] *See infra* note 165 and accompanying text.
[27] *See infra* note 167 and accompanying text.

I. THE VARIED OBJECTIVES OF STATUTORY INTERPRETATION

Understanding how data-driven interpretive sources fit into the structure of statutory interpretation requires consideration of two basic components of textual interpretation: (1) the *constituent question* of the proper objective of interpretation, and (2) the *evidential question* of which sources of meaning help determine the standard set by the constituent question.[28] Commentators typically answer the constituent question by advocating that courts should seek to determine the beliefs or actions of some particular classification of people. Sweeping broadly, the various forms of "intentionalism" focus on the language production of the legislature while the various forms of "textualism" focus on language comprehension, typically of the ordinary person or interpretive community.[29] Notwithstanding differing interpretive perspectives, courts agree that, to some degree at least, language comprehension should be prioritized. Courts thus presume that language in legal texts should be given its "ordinary meaning," determined by general principles of language usage that apply outside the law.[30] The ordinary meaning standard is justified in part on the basis that it is consistent with fundamental principles of legal interpretation, such as notice, predictability, and the notion that the public should be able to read and understand legal texts.[31]

---

[28] *See* BRIAN G. SLOCUM, ORDINARY MEANING: A THEORY OF THE MOST FUNDAMENTAL PRINCIPLE OF LEGAL INTERPRETATION 36–37 (2015) (describing the constituent and evidential questions).

[29] This essay distinguishes between language production, as measured by the legislature at issue, and language comprehension, as measured by ordinary people, but interpretive theory could also focus on other speech communities. *See* William S. Blatt, *Interpretive Communities: The Missing Element in Statutory Interpretation*, 95 Nw. U. L. Rev. 629, 630–31 (2001) (explaining that "[s]tatutes engage the following three distinct communities: the policy community of specialized professionals found in government bureaucracies, the political community of elected politicians, and the public community of the general electorate"). Nevertheless, this essay focuses on interpretive sources, rather than speech communities, and basic (but overlooked) distinctions amongst speech communities are sufficient to illustrate the important differences among interpretive sources.

[30] *See* SLOCUM, *supra* note 28, at 1–3.

[31] *See* Herman Cappelen, *Semantics and Pragmatics: Some Central Issues*, *in* CONTEXT-SENSITIVITY AND SEMANTIC MINIMALISM: NEW ESSAYS ON SEMANTICS AND PRAGMATICS, 3, 18–19 (Gerhard Preyer & Georg Peter eds., 2007) (explaining that "[w]hen we articulate rules, directives, laws and other action-guiding instructions, we assume that people, variously situated, can grasp that content in the same way").

A. Ordinary Meaning as the Objective of Statutory Interpretation

Although courts reflexively cite to the ordinary meaning doctrine and offer clues to its features, they have not defined the principle, nor do they use it consistently.[32] Even basic questions like the proper focus of the ordinary meaning interpretation (whether individual words or something larger such as the sentence) are undertheorized and subject to inconsistent treatment by courts.[33] Some aspects of the ordinary meaning doctrine are thus contestable, but by its very nature the "ordinary" meaning of a provision must consist of elements that cut across contexts (different statutes, subjects, congresses, etc.).[34] Otherwise, the "ordinary" meaning concept would be incoherent because it would be based only on the specific context of the statute and perceived intent of the legislature.[35] Consider, for instance, a highly unusual meaning that a court nevertheless deems to be the intended meaning based on consideration of the entire context of a statute. While that meaning might be said to be the "correct" meaning, there would be nothing "ordinary" about such a meaning.[36] Ordinary meaning might thus be defined in something like the following way: *the meaning the language conveys to an ordinary member of the community based on conventions of meaning and other systematicities of language, such as compositionality, and a notion of context that helps select the appropriate conventional meanings of the relevant terms*.[37] As the description indicates, the ordinary meaning concept focuses on "what the ordinary reader of a statute would have understood the words to mean at the time of enactment," as opposed to the legislature's intent in creating the statute.[38]

---

[32] *See generally* LAWRENCE M. SOLAN, THE LANGUAGE OF STATUTES: LAWS AND THEIR INTERPRETATION (John M. Conley & Lynn Mather eds., 2010).

[33] *See* SLOCUM, *supra* note 28, at 106 (explaining that "[p]art of the problem with the current judicial approach to interpretation . . . is that courts often frame the ordinary meaning inquiry as involving an individual word instead of the relevant sentence").

[34] *See* Brian G. Slocum & Jarrod Wong, *The Vienna Convention and the Ordinary Meaning of International Law*, 46 YALE J. INT'L L. (forthcoming 2021) (manuscript at 8).

[35] This description is somewhat simplified because even if an interpreter focuses on the specific context of a statute and perceived intent of the legislature, the interpreter cannot make sense of the language without relying on conventions of language. *See* SLOCUM, *supra* note 28, at 54–72.

[36] *See* Slocum & Wong, *supra* note 34, at 35.

[37] *See* SLOCUM, *supra* note 28, at 57 (describing the importance of conventions to meaning). "The principle of compositionality states that the meaning of a complex linguistic expression is built up from the meanings of its composite parts in a rule-governed fashion." M. LYNNE MURPHY & ANU KOSKELA, KEY TERMS IN SEMANTICS 36 (2010). Thus, a sentence is compositional if its meaning is the sum of the meanings of its parts and of the relations of the parts.

[38] Thomas W. Merrill, *Textualism and the Future of the* Chevron *Doctrine*, 72 WASH. U. L.Q. 351, 351–52 (1994). Certain Supreme Court opinions also focus on the likely

Despite the focus on conventions of meaning, consideration of context and purpose are ineliminable aspects of the ordinary meaning determination. With natural-language understanding, and particularly with legal texts, the objective is to determine what a sentence means in a given context of utterance rather than just what it could mean in general.[39] An ordinary meaning must therefore be based on conventions of language, rather than inferences about speaker intent, but also upon consideration of some kinds of contextual and purposive evidence.[40] Still, when a court seeks to determine the ordinary meaning of a text, inferences from context should be relevant to conventions of language, rather than some broader notion of authorial intent.[41]

Despite its importance, as defined above, ordinary meaning cannot serve as the single answer to statutory interpretation's constituent question because it often underdetermines the actual interpretations made by courts.[42] For instance, a court may rely on legal concerns that are in tension with the ordinary meaning of the textual language, and as a result the ordinary meaning will not coincide with the meaning the court gives the provision.[43] Alternatively, it might be clear that the relevant textual language should be given a special legal or technical meaning, or even some meaning that is not technical or legal but is seldom used (and thus an unordinary meaning).[44] Furthermore, an ordinary meaning may be general or vague and therefore indeterminate in light of the precise, yes/no distinctions often required to resolve interpretive disputes.[45] In the above situations, and often as a general matter, the court might rely on contextual inferences about the meaning

---

interpretation or an ordinary person. *See, e.g.*, Bond v. United States, 572 U.S. 844, 861 (2014) (explaining that "[w]hen used in the manner here, the chemicals in this case are not of the sort that an ordinary person would associate with instruments of chemical warfare").

[39] *See* SLOCUM, *supra* note 28, at 111–12 (discussing the importance of context to interpretation).

[40] *See id.*

[41] *See* Slocum & Wong, *supra* note 34, at 24 (describing how the same interpretive tool can be used to select the appropriate conventional meaning or, instead, be used as a basis for broad inferences about legislative intent).

[42] The extent to which ordinary meaning underdetermines a court's interpretation depends, obviously, on how broadly "ordinary meaning" is defined. While a very narrow definition of ordinary meaning may be unsatisfactory because it underdetermines interpretations in every case, an unduly broad definition will lead to incoherence because it serves merely as a conclusory label for whatever interpretation a court finds to be most persuasive.

[43] *See infra* notes 47–51 and accompanying text.

[44] Thus ordinary meaning is defeasible. *See, e.g.*, Taniguchi v. Kan Pac. Saipan, Ltd., 566 U.S. 560, 569 (2012) (explaining that "the word 'interpreter' can encompass persons who translate documents, but because that is not the ordinary meaning of the word, it does not control unless the context in which the word appears indicates that it does").

[45] *See* Brian G. Slocum, *Replacing the Flawed* Chevron *Standard*, 60 WM. & MARY L. REV. 195, 238–39 (2018) (explaining how legal interpretation "in general relies on bivalency (i.e., the idea that interpretative questions have 'yes' or 'no' answers)," which is in tension with the prototypical structure of language).

communicated by the text. The inferences may be in tension with, or more precise than, applicable conventions of language.[46]

B. Communicative Meaning and Legal Meaning as Alternative Objectives of Statutory Interpretation

If a court goes beyond ordinary meaning, as described above, and seeks the meaning communicated by a provision, it must choose some perspective from which to base the determination. A court could orient the perspective to the reader and attempt to determine the text's *reader comprehension communicative meaning*, which may be understood as *the meaning an appropriate reader would most reasonably take an author*[47] *to be trying to convey in employing a given verbal vehicle in the given communicative-context*.[48] Alternatively, a court could orient the perspective to the author and attempt to determine the text's *language production communicative meaning*, which may be understood as *the meaning the legislature was trying to convey in employing a given verbal vehicle in the given communicative-context*.[49] Conceivably, the *evidential question* of which sources of meaning help determine the standard set by the constituent question could differ depending on which orientation the court chooses.[50]

However it is framed, the communicative meaning of a provision often underdetermines a court's ultimate interpretation. For example, the communicative meaning of the textual language

---

[46] For example, the lexical meaning of "vehicle" may be indeterminate regarding certain items that are not prototypical vehicles but also not clearly nonvehicles. Hart, in his famous no-vehicles-in-the-park hypo, indicated that it might be unclear whether "vehicle" includes such things as bicycles. *See* H.L.A. Hart, *Positivism and the Separation of Law and Morals*, 71 HARV. L. REV. 593, 607 (1958). It may be clear from the statutory scheme, however, that a bicycle should be considered a vehicle, or that some other marginal item should not be considered a vehicle. *See, e.g.*, State v. Barnes, 403 P.3d 72, 75–76 (2017) (holding that a riding lawn mower was not a "motor vehicle" within the meaning of a theft of a motor vehicle statute because the statute was enacted to combat the high rate of automobile theft and associated crimes and not the theft of lawn mowers).

[47] For ease of writing, the singular "author" is used even though legislatures are multi-member bodies (whose members generally are not the actual draftors of the statutory text).

[48] It could be debated whether the definition should be framed in terms of what the author was "trying to convey," as opposed to what was successfully conveyed. *See* SLOCUM, *supra* note 28, at 70 (discussing whether "successful communication" should be an element of any definition of meaning). For purposes of this article, however, resolution of the issue is not necessary. It is sufficient to realize that the meaning attempted (whether successful or not) by the legislature may differ from the meaning understood by ordinary readers.

[49] As the definition suggestions, the determination of *language production communicative meaning* does not rely on the language comprehension of ordinary readers but, rather, on the processes by which Congress enacts legislation. *See, e.g.*, Gluck, *supra* note 1, at 80–96 (analyzing statutory interpretation through the processes used by Congress in enacting legislation).

[50] *See infra* notes 57–60 and accompanying text.

may be legally unacceptable for some reason, such as a meaning that would raise a serious constitutional issue or result in absurdity.[51] In such situations, the court's interpretation is based on principles that reflect normative legal commitments rather than language production or comprehension.[52] The court is thus determining the *legal meaning* of the provision, which may be understood as something like *the meaning a competent judge would give to the text in light of its ordinary or communicative meaning as modified by concerns specific to the law*.[53]

C. The Importance of Identifying the Objective of Statutory Interpretation

The above description offers a broad, schematic overview of the basic elements of statutory interpretation and omits various nuances not relevant to the ideas developed in this essay. For instance, the three categories outlined above are contestable, others might organize or define the typical interpretive goals differently, and only "ordinary meaning" is a term that courts currently use when describing their approach to statutory interpretation.[54] Nevertheless, for a few reasons, it is useful to think about the varied interpretive goals of courts, even at a high level of generality.

First, the meaning a court gives a statute is not always synonymous with the ordinary meanings of its terms.[55] Thus, descriptive coherence requires additional explanatory concepts to

---

[51] The absurdity doctrine may be the clearest example of a situation where a court has rejected the meaning of the text (whether communicative or otherwise) in favor of some other meaning. *See* John F. Manning, *The Absurdity Doctrine*, 116 HARV. L. REV. 2387, 2388 (2005) (describing how courts have long embraced the idea that "judges may deviate from even the clearest statutory text when a given application would otherwise produce 'absurd' results"). The avoidance canon is another example where the interpretation chosen by the court might not conform to the intended meaning of the statute or the meaning an ordinary reader would give it. *See, e.g.*, Eric S. Fish, *Constitutional Avoidance as Interpretation and as Remedy*, 114 MICH. L. REV. 1275, 1275 (2016) (arguing that the Court has used the avoidance canon to "rewrite laws").

[52] *See* Fish, *supra* note 51.

[53] Gaps between communicative meaning and legal meaning can also arise when an interpretive source's definition has changed and a court applies the principle to legislation enacted prior to the change. *See, e.g.*, William S. Dodge, *The New Presumption Against Extraterritoriality*, 133 HARV. L. REV. 1582, 1589 (2020) (noting that "over time, the presumption against extraterritoriality has changed significantly," and has evolved "from a rule based on international law, to a canon of comity, to an approach for determining legislative intent").

[54] One could, for instance, make many more distinctions amongst the different varieties of meaning. *See* Richard H. Fallon, Jr., *The Meaning of Legal "Meaning" and Its Implications for Theories of Legal Interpretation*, 82 U. CHI. L. REV. 1235, 1243–52 (2015). While doing so may be useful for some purposes, the objective of this Article is to illustrate the differences between the basic perspectives of production and comprehension and how big data may be relevant to one perspective but not the other.

[55] *See supra* notes 40–43 and accompanying text.

depict the nature of statutory interpretation.[56] Second, considering the varied interpretive goals of courts allows for a more precise and nuanced analysis of the sometimes shifting ways in which judges focus on language production or language comprehension. The ordinary meaning doctrine's orientation is language comprehension, but when a provision's ordinary meaning underdetermines the meaning chosen by the court, the court's orientation must have shifted (even if implicitly) during the interpretive process.[57] The court might have determined the provision's communicative meaning either by focusing on language production or language comprehension.[58] Alternatively, the court might have instead focused on legal meaning, which may be difficult to tie to either language production or language comprehension.[59] Whatever the interpretive objective, the court can then select interpretive sources relevant to that objective.

Finally, when a court discusses an interpretive source, it is useful to consider the judiciary's fluctuating interpretive orientations in assessing whether the court is employing a legal fiction.[60] It may be that an interpretive source relating to language production is a legal fiction if it is assumed that an ordinary person would consult it.[61] Similarly, an interpretive source relating to language comprehension may be a legal fiction if it assumed that the legislature adheres to it when drafting.[62] Of course, courts often attempt to legitimize interpretive sources by tying them to both language production and language comprehension. For instance, the very premise of the ordinary meaning doctrine is that the test for meaning is an objective one that is external to the legislature's

---

[56] However the objectives of interpretation are defined and organized, it is clear that no single concept is sufficient to explain the interpretive process.

[57] Of course, it may be that some form of communicative meaning was the court's interpretive objective from the beginning of the interpretive process. Even if communicative meaning is the court's interpretive objective, it is useful to consider whether the court is focusing on language comprehension or language production, as well as whether the court rejects the communicative meaning based on legal principles.

[58] *See supra* notes 47–49 and accompanying text.

[59] It is possible to fit at least some substantive canons within the concept of communicative meaning, but doing so likely involves legal fictions. *See generally* Brian G. Slocum, *Reforming the Canon of Constitutional Avoidance*, U. PA. J. CONST. L. (forthcoming 2021) (providing a theory of how the canon of constitutional avoidance can be viewed as an aspect of communicative meaning).

[60] A legal fiction is "an assumption of fact deliberately, lawfully and irrebutably made contrary to the facts proven or probable in a particular case, with the object of bringing a particular legal rule into operation or explaining a legal rule, the assumption being permitted by law or employed in legal science." Todd Barnet, *Legal Fiction and Forfeiture: An Historical Analysis of the Civil Asset Forfeiture Reform Act*, 40 DUQ. L. REV. 77, 80 (2001).

[61] *See infra* notes 82–84 and accompanying text (discussing the *in pari materia* canon).

[62] *See* Gluck & Bressman, *supra* note 7, at 911–24 (analyzing whether legislative drafters consider various interpretive principles when drafting legislation). The same interpretive principle may still be a fiction when it is assumed that it is relevant to the language comprehension of ordinary people.

actual intentions.[63] Thus, the focus is on the language comprehension of ordinary people. Still, a court can simply assert that the ordinary meaning standard is an aspect of language production. In fact, according to the Supreme Court, courts "assume that the ordinary meaning of [the statutory] language accurately expresses the legislative purpose."[64] Of course, this assumed intent is generalized in the sense that it is not connected to any particular Congress, subject matter, or statute. In short, the assumption is contestable in many situations.

II. INTERPRETIVE SOURCES AND THE OBJECTIVES OF INTERPRETATION

A chosen objective of interpretation must therefore provide an answer to the constituent question of interpretation, and an interpretive source, when relevant, should provide evidence that helps select the meaning that best satisfies the objective of interpretation. A coherent interpretive process will include sources of meaning that, when applied accurately, select a meaning defined by the objective of interpretation and exclude sources that do not do so.[65] Table 1 below lists the three objectives described above, along with some of the interpretive sources that provide information relevant to the given objective.

---

[63] *See supra* notes 39–40 and accompanying text.

[64] Marx v. Gen. Revenue Corp., 568 U.S. 371, 376 (2013) (alteration in original) (citing Hardt v. Reliance Standard Life Ins. Co., 560 U.S. 242, 251 (2010)); *see also* Am. Tobacco Co. v. Patterson, 456 U.S. 63, 68 (1982) (explaining that the Court "assume[s] that the legislative purpose is expressed by the ordinary meaning of the words used" (internal quotations omitted)).

[65] That is, the sources of meaning, and how they can be used, should be consistent with the requirements of the objective of interpretation.

Table 1

Three Objectives of Interpretation with Corresponding Evidential Sources

Ordinary Meaning —word meaning (dictionaries, intuition, precedent) —sentential (and maybe broader) context —grammatical rules Communicative Meaning (language production and comprehension) —legislative history —*in pari materia* Legal Meaning —absurdity doctrine —substantive canons

A. Interpretive Sources Relevant to Ordinary Meaning

Consider the ordinary meaning objective. Various interpretive sources arguably provide information relevant to the determination of *the meaning the language conveys to an ordinary member of the community based on conventions of meaning*.[66] Dictionaries are an obvious example.[67] Courts consult dictionary definitions based on the (often mistaken) belief that a definition reflects general usage by the public, and thus the ordinary meaning of the word in the statute, as opposed to a belief that the definition reveals some particular legislative intent.[68] Similarly, grammatical rules provide information about the ordinary meaning of a provision, as long as those rules accurately depict how the given language community interprets language. Thus, for example, a rule about a comma is relevant to the ordinary meaning of a text if an ordinary member of the community would consider it (even implicitly) when interpreting a provision.[69]

---

[66] *See* Slocum & Wong, *supra* note 34, at 67–76.

[67] To be sure, judicial reliance on dictionaries has been harshly criticized by commentators. *See, e.g.*, Ellen P. Aprill, *The Law of the Word: Dictionary Shopping in the Supreme Court*, 30 ARIZ. ST. L.J. 275, 277, 280–81 (1998) (describing the unprincipled use of dictionaries by judges).

[68] *See, e.g.*, Utah v. Evans, 536 U.S. 452, 475 (2002) (purporting to get information about "[c]ontemporaneous general usage" about the meaning of "enumeration" from dictionary definitions).

[69] Thus, if a rule that indicates statutory meaning from the placement of a comma does not reflect general usage, it cannot be an interpretive source for the determination of ordinary meaning (and thus may have no legitimate use). *See* Lance Phillip Timbreza, *The Elusive Comma:*

The surrounding linguistic context of a provision can also provide information relevant to the determination of ordinary meaning. In fact, contextual considerations are such an integral aspect of meaning that even interpretive sources based on generalized intent and systematicities of language usage may require consideration of the particularized context of the statute.[70] For instance, ordinary meaning likely includes at least some textual canons, which are varied presumptions about meaning "that are drawn from the drafter's choice of words, their grammatical placement in sentences, and their relationship to other parts of the 'whole' statute."[71] The presumptions typically are said to be based on general principles of language usage rather than legal concerns.[72] Importantly, though, textual canons, to varying degrees, require courts to consider the context of the statute, making the systematicity of language identified by the textual canon only one aspect of its application.[73] Thus, the interpretive canon may be justified by its consistency with general linguistic usage, but the actual application of the canon may call for consideration of the particularized context of the statute (which may even convince the court that the general linguistic usage that the canon represents should not apply).[74]

B. Interpretive Sources Relevant to Communicative Meaning

Conversely, other interpretive sources may provide information relevant to one framing of communicative meaning but not another. Consider, for example, legislative history. Legislative history provides information relating to the language production of the legislature, not the conventions of meaning applied by an

---

The Proper Role of Punctuation in Statutory Interpretation, 24 QLR 63, 67 (2005) (explaining the Supreme Court's creation of "Punctuation Doctrines" for statutory interpretation).

[70] *See* Anita S. Krishnakumar, *Backdoor Purposivism*, 69 DUKE L.J. 1275, 1331 (2020) (explaining how the application of textual canons allows courts to engage in purposivist reasoning).

[71] *See* WILLIAM N. ESKRIDGE & PHILIP P. FRICKEY, CASES AND MATERIALS ON LEGISLATION: STATUTES AND THE CREATION OF PUBLIC POLICY 634 (1995); *see also* SLOCUM, *supra* note 28, at 181–212 (analyzing whether various determinants of meaning fall under the ordinary meaning doctrine).

[72] *See* Abbe R. Gluck & Richard A. Posner, *Statutory Interpretation on the Bench: A Survey of Forty-Two Judges on the Federal Courts of Appeals*, 131 HARV. L. REV. 1298, 1330 (2018) (distinguishing between "'linguistic' or 'textual' canons, which are presumptions about how language is used," and "normative" or "policy" canons).

[73] *See* Krishnakumar, *supra* note 70, at 1291 (arguing that some judges use textual canons in broad, purposivist ways that serve as "launch pads for assuming or constructing legislative purpose and intent" (emphasis omitted)).

[74] *See id.*

ordinary member of the community.[75] Legislative history is therefore relevant to *language production communicative meaning*, but its relevance to *reader comprehension communicative meaning* is not clear.[76] Compared to ordinary meaning, a reader-based framing of communicative meaning allows for greater consideration of contextual evidence because any contextual evidence is not limited to that which *helps select the appropriate conventional meaning of the relevant terms*.[77] The difficult question becomes which sources of meaning an "appropriate reader" would consider. Is this an empirical question or a normative one?[78] If empirical, and an "appropriate reader" is synonymous with an average member of the community, it may require a legal fiction for a court to consult legislative history when determining *reader comprehension communicative meaning*.[79]

Other interpretive sources, such as some textual canons, may be more difficult to categorize. Courts and commentators typically legitimize textual canons as reflecting general principles of language usage rather than legal concerns.[80] The key question, though, is whether textual canons measure the language comprehension of ordinary members of the community or, conversely, the language usage of the legislature. It may be that they do not measure either accurately.[81] Consider the *in pari materia* canon, which presumes that the same word in two related provisions of a statute will have the same meaning, and also assumes other forms of internal statutory coherence.[82] Are these presumptions relevant to language comprehension?

---

[75] *See* Stuart Minor Benjamin & Kristen M. Renberg, *The Paradoxical Impact of Scalia's Campaign Against Legislative History*, 105 CORNELL L. REV. 1023, 1025–28 (2020) (discussing arguments about whether legislative history accurately reveals legislative intent). *Cf.* Gluck & Posner, *supra* note 72, at 1336–37 (surveying judges about their interest in information about the legislative process and creation of statutes).

[76] *See supra* note 47 and accompanying text (defining *reader comprehension communicative* meaning).

[77] *See supra* note 37 and accompanying text (defining *ordinary meaning*).

[78] To the extent courts desire to view the question as empirical, there is currently no evidence about the interpretive sources an average member of the community would consider. *Cf.* Christopher Brett Jaeger, *The Empirical Reasonable Person*, 72 ALA. L. REV. 887, 889 (2021) (explaining that the reasonable person standard has "always been more philosophical than empirical" and very little empirical work exists regarding how laypeople make the determination of how a reasonable person would act).

[79] *See supra* note 60 and accompanying text (explaining legal fictions).

[80] *See supra* notes 71–73 and accompanying text (describing textual canons).

[81] *See generally* Kevin Tobia, Brian G. Slocum & Victoria Nourse, *Statutory Interpretation from the Outside*, COLUM. L. REV. (forthcoming 2022) (using surveys to determine whether ordinary people implicitly apply various interpretive canons).

[82] *See* Jacob Scott, *Codified Canons and the Common Law of Interpretation*, 98 GEO. L.J. 341, 376 (2010) (explaining that "[t]he presumption of consistent usage and *in pari materia*, which both accept an interpreter's examination of the context of a particular term and what sort of meaning that term has acquired in other statutes, are implicitly the same canon as the presumption of consistency between statutes").

Applying the *in pari materia* canon typically requires an in-depth knowledge of the legal system, which an ordinary member of the community would not possess.[83] Thus, linking *in pari materia* to the language comprehension of an ordinary person would likely involve fictional assumptions about the ability of readers to gather the information relevant to the application of the canon and to make the required assessments of coherence and consistency, as well as their interest in doing so. The concerns may be equally fictional if the focus is on language production. It may be that legislative drafters do not consider *in pari materia* concerns when drafting, or are limited by the legislative process in doing so.[84] Instead, *in pari materia* might be a normative concept that reflects judicial values.

C. Interpretive Sources Relevant to Legal Meaning

When interpretive sources are relevant to *legal meaning*, classification difficulties are even more pronounced.[85] Consider substantive canons, also referred to as "normative canons," among other terms,[86] which are "presumptions about statutory meaning based upon substantive principles or policies drawn from the common law, other statutes, or the Constitution."[87] The strongest substantive canons are "clear statement rules" and require a court to avoid a particular result unless the statute (more clearly than usually required) indicates that the result was intended.[88] For instance, the presumption against retroactivity directs courts to select a prospective-only interpretation unless the language of the provision clearly indicates that the legislature intended for the

---

[83] For instance, in determining the meaning of the stipulated definition of "take" in the Endangered Species Act of 1973, an ordinary person may not infer from a separate provision providing for permits for takings that a broad meaning of "take" was congressionally intended. *See* Babbitt v. Sweet Home Chapter of Cmtys. for a Great Oregon, 515 U.S. 687, 700–01 (1995) (making such an inference).

[84] *See* Gluck & Bressman, *supra* note 7, at 1021 n.455 (finding based on surveys that "those who attended elite schools were more likely to know *in pari materia*").

[85] *See supra* note 53 and accompanying text (defining legal meaning).

[86] *See* Stephen F. Ross, *Where Have You Gone, Karl Llewellyn? Should Congress Turn its Lonely Eyes to You?*, 45 VAND. L. REV. 561, 563 (1992) (referring to substantive canons as "normative canons").

[87] *See* ESKRIDGE & FRICKEY, *supra* note 71, at 634.

[88] *See Clear Statement Rules, Federalism, and Congressional Regulation of States*, Note, 107 HARV. L. REV. 1959, 1959 (1994) (noting that clear statement rules "erect potential barriers to the straightforward effectuation of legislative intent"); *see also* William N. Eskridge, Jr., & Philip P. Frickey, *Quasi-Constitutional Law: Clear Statement Rules as Constitutional Lawmaking*, 45 VAND. L. REV. 593, 598 (1992) (arguing that the Court's clear statement rules "amount to a 'backdoor' version of the constitutional activism that most Justices on the current Court have publicly denounced").

statute to apply retroactively.[89] Thus, the presumption against retroactivity allows courts to "infer exceptions [creating prospective-only applications] to statutory provisions whose words, on their face, appear to cover all pending cases."[90] The Court has asserted that "[b]ecause it accords with widely held intuitions about how statutes ordinarily operate, a presumption against retroactivity will generally coincide with legislative and public expectations."[91]

The Court's statement raises the obvious question of how it is aware of "public expectations" regarding "how statutes ordinarily operate."[92] Certainly, in terms of specific applications of the canon, the presumption against retroactivity requires knowledge that might exceed that of an ordinary member of the community. An interpreter applying the canon must understand the legal system and the concept and definition of retroactivity, as well as the notion that the literal meaning of statutory language is not always synonymous with its legal meaning.[93] Perhaps, however, the canon reflects a very general public assumption about how the law operates, even if an ordinary member of the public would be confused about its application in any given case.[94]

The conventional wisdom is that substantive canons are based on normative concerns specific to the law, and undoubtedly courts interpret statutes in light of these sorts of legal concerns.[95] Nevertheless, the Court's assertion that the presumption against retroactivity is consistent with public expectations illustrates the judicial impulse to connect all interpretive sources to the language production of Congress or the language comprehension of some relevant language community. Data-driven interpretive sources like corpus linguistics and surveys must therefore be similarly linked to some objective of interpretation relating to either the language production of Congress or the language comprehension of some speech community (such as ordinary people). Furthermore,

---

[89] I.N.S. v. St. Cyr, 533 U.S. 289, 316–17 (2001) (explaining that a statute must be "so clear that it could sustain only one interpretation" before it will be given retroactive effect).

[90] *See* Caleb Nelson, *What is Textualism?*, 91 VA. L. REV. 347, 384 (2005).

[91] Landgraf v. USI Film Prods., 511 U.S. 244, 272 (1994); *see also* Ronald M. Levin, *"Vacation" at Sea: Judicial Remedies and Equitable Discretion in Administrative Law*, 53 DUKE L.J. 291, 349 (2003) (noting that the Court's motivation for the presumption against retroactivity is the unfairness involved in retroactive legislation and concern for the rule of law).

[92] *Landgraf*, 511 U.S. at 272. For that matter, the Court's statement also raises the question of how it is aware of legislative expectations.

[93] For instance, "deciding when a statute operates 'retroactively' is not always a simple or mechanical task," but rather requires that a "court must ask whether the new provision attaches new legal consequences to events completed before its enactment." *Landgraf*, 511 U.S. at 268–70.

[94] For instance, it may be that the public believes that statutes should, when possible, be interpreted in a manner consistent with "fundamental national principles." *See* Einer Elhauge, *Preference-Eliciting Statutory Default Rules*, 102 COLUM. L. REV. 2162 , 2256 (2002).

[95] *See supra* notes 48–50 and accompanying text.

to be legitimate, they must serve as reliable sources that help select the meaning that best satisfies the objective of interpretation.

III. CORPUS LINGUISTICS WITHIN THE STRUCTURE OF INTERPRETATION

Parts I and II offer a framework for conceptualizing the role of corpus linguistics within statutory interpretation, including whether corpus analysis might change how claims of interpretive accuracy are evaluated. In a basic sense, it is relatively simple to fit corpus linguistics within the structure of interpretation. As currently practiced, corpus analysis is legitimized (if at all) through its connection to the ordinary meaning doctrine and therefore can be expected to provide information relevant to accepted and typical standards of communication that apply outside of the law.[96] Primarily, corpus linguistics can provide information about lexical meaning that, arguably, is superior to some widely used interpretive sources, such as dictionaries.[97] Corpus linguistics can illustrate the "number of senses (i.e., meanings) a linguistic expression may have" and its most frequently used meaning.[98] It can also provide information about the "most prototypical meaning of an expression," based on various factors.[99]

While the role of corpus linguistics as an ordinary meaning interpretive source should seem rather straightforward, as explained below, both proponents and opponents of corpus linguistics have framed it in ways that exaggerate its potential. Certainly, various aspects of corpus linguistics are contestable, such as whether corpus linguistics provides accurate information about the ordinary meanings of words and whether the analysis is too difficult for judges to perform competently.[100] This Part, though, focuses only on issues concerning how corpus linguistics fits within the structure of interpretation and why corpus linguistics cannot transform statutory interpretation into an empirical issue. Corpus linguistics can reveal important systematicities of language usage, but the information it provides is limited in important ways. Corpus linguistics does not represent a theory of legal interpretation (contrary to what some opponents have claimed), and arguments (by some proponents) that corpus linguistics can help discern legislative

---

[96] *See* Gries & Slocum, *supra* note 22, at 1422–33.

[97] *See id.* at 1441–42 (explaining that corpus analysis is superior to dictionaries because corpus analysis can take account of statutory context in ways that dictionaries cannot).

[98] *Id.* at 1441.

[99] *Id.*

[100] *See* Kevin Tobia, *The Corpus and the Courts*, 3/5/2021 U. CHI. L. REV. ONLINE 1 (discussing criticisms of corpus linguistics and offering a set of best practices for its use within legal interpretation).

intent or transform statutory interpretation into an empirical science fail to appreciate the intensely contextual nature of statutory interpretation.

A. Arguments by Critics that Exaggerate the Role of Corpus Linguistics

It may be tempting for a critic to label a new interpretive source as a new interpretive theory. After all, as Parts I and II explained, interpretive theories implicate the judicial function and are basic and fundamental aspects of legal interpretation, while interpretive sources play a subservient role in support of interpretive theories. An example of such labeling is Carissa Hessick's argument that "[i]t is easy to overlook that corpus linguistics is an interpretive theory, rather than simply an interdisciplinary methodology, because it bills itself as providing an answer to a question that many current interpretive theories ask: What is the 'plain' or 'ordinary' meaning of the statutory text?"[101] The very language of Hessic's assertion undermines her claim that corpus linguistics is an "interpretive theory." An interpretive theory must, at a minimum, purport to answer the constituent question of interpretation.[102] Thus, for example, *language production communicative meaning* is an interpretive theory because it sets forth a standard or objective of interpretation, namely that a court should seek to determine *the meaning the legislature was trying to convey in employing a given verbal vehicle in the given communicative-context*.[103] In contrast, corpus linguistics does not purport to answer the constituent question of legal interpretation. Rather, it provides information relevant to an objective of interpretation posited by some actual theory of interpretation.[104]

Corpus linguistics is no more a theory of legal interpretation than Westlaw is a theory of legal interpretation. Certainly, in some general sense, any interpretive source is based on a theory about language. The assumption underlying most corpus-based analyses is the so-called distributional hypothesis, which provides that words that are used in and occur in the same

---

[101] Carissa Byrne Hessick, *Corpus Linguistics and the Criminal Law*, 2017 BYU L. REV. 1503, 1505 (2017).

[102] *See supra* notes 28–29 and accompanying text (describing the constituent question of interpretation).

[103] *See supra* note 49 and accompanying text (describing language production communicative meaning).

[104] *See* Brian G. Slocum & Stefan Th. Gries, *Judging Corpus Linguistics*, 94 S. CAL. L. REV. POSTSCRIPT 13, 19–20 (2020).

contexts have similar meanings.[105] Currently, the contexts used for corpus linguistics data gathering involve language produced by people other than Congress.[106] As such, corpus linguistics provides information about how language is likely to be understood by ordinary readers (or some other language community), rather than how language is produced by Congress.[107] A corpus analysis is thus more useful for quantifying to what degree a certain intention is encoded in a text in such a way that it will be understood by ordinary readers than for inferring the intentions of the producers of the text.[108] Thus, in a narrow sense, corpus linguistics is based on a theory of language meaning, but it does not purport to provide any answers to legal questions involving the proper judicial function or objective of interpretation.[109] It is therefore not a theory of legal interpretation, and is a source of meaning only to the extent that language comprehension is relevant to the interpreter.

As an interpretive source that provides information relevant to the ordinary meaning of textual language, corpus linguistics can further the long-standing judicial practice of legitimizing interpretations by distancing them from the personal predilections of judges.[110] The original impetus behind such efforts may have been a desire to deflect accusations of ideologically motivated reasoning, but with the increasing prevalence of big data interpretive sources, an equally powerful motivation is the selection of interpretive sources that accurately determine linguistic meaning.[111] A requirement to understanding this motivation, though, involves acknowledging the differences between individual and collective usages of language. Lexical meaning is based on collective usage and not the views of any one person.[112] If individual intuitions about word meanings were always accurate, there would be no need for judges to consider evidence external to themselves about

---

[105] Stefan Th. Gries, *What is Corpus Linguistics?*, 3 LANGUAGE AND LINGUISTICS COMPASS, 1225, 1226–28 (2009).

[106] *See id.* at 1229–32.

[107] *See id.*

[108] *See id.*

[109] As a methodology from linguistics, corpus linguistics is naturally based on linguistic principles rather than legal ones. *See* Slocum & Gries, *supra* note 104, at 13–14, 30–31.

[110] Legal interpretation has long sought to offer an objective view of the interpretive process that distances an interpretation from the interpreter. *See, e.g.*, Oliver Wendell Holmes, *The Theory of Legal Interpretation*, 12 HARV. L. REV. 417, 417–18 (1899) (indicating that the interpreter's role is to determine "what [] words would mean in the mouth of a normal speaker of English, using them in the circumstances in which they were used").

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

[112] *See* RONNIE CANN, RUTH KEMPSON, & ELENI GREGOROMICHELAKI, SEMANTICS: AN INTRODUCTION TO MEANING IN LANGUAGE 4–7 (2009) (explaining how researchers can construct meanings based on how words are used in society).

lexical meaning. Scholars have known for some time, however, that individual intuitions about lexical meaning are often mistaken.[113] Making matters worse, individuals overestimate the extent to which their views about meaning correspond with how most other people view the same issues.[114]

If corpus linguistics provides evidence about collective usage, and lexical meaning is based on collective usage, criticisms of corpus linguistics should focus on such issues as whether the collective usage measured by a corpus analysis is appropriate to the interpretation of a legal text or whether corpus analysis is accurate in measuring that usage. An example of the former is an argument that legal interpretation should focus on the language production of the legislature, but existing corpora contain only files from non-legal sources.[115] Thus, if the proper focus of statutory interpretation is language production, corpus analysis would not provide useful information. An example of the latter argument is the claim that corpus analysis is inaccurate because it produces data that is relevant to prototypical meaning rather than the extent of possible meanings.[116] Thus, corpus analysis might systematically provide information that would cause words to be defined too narrowly.

Arguments that conflate individual intuitions about language meaning with evidence about collective usage are therefore misplaced. Consider one such example: Hessick notes that a premise of corpus linguistics is that judicial intuition regarding lexical meaning "ought to be replaced with corpus analyses precisely because that intuition is unreliable."[117] Hessick reasons that if corpus linguistics results "will always be the same as intuition, then corpus linguistics is unnecessary. If there are cases where corpus linguistics returns a different result than would the intuitions of lawmakers, judges, or voters, then corpus linguistics is a real threat to notice and accountability."[118] Hessick, essentially, presents a false dilemma. Recall that ordinary meaning is based on collective usage, and corpus linguistics provides evidence of collective usage.[119] Any individual's intuitions (whether lawmaker, judge, or voter) may

---

[113] *See* Lawrence Solan et al., *False Consensus Bias in Contract Interpretation*, 108 COLUM. L. REV. 1268, 1268 (2008) (explaining the concept of "false consensus bias," which describes the propensity to believe that one's views about meaning are the predominant views).

[114] *See id.* at 1269.

[115] *See* Anya Bernstein, *Legal Corpus Linguistics and the Half Empirical Attitude*, 106 CORNELL L. REV. (forthcoming 2021) (manuscript at 3).

[116] *See* Kevin P. Tobia, *Testing Ordinary Meaning*, 134 HARV. L. REV. 726, 734–35 (2020).

[117] *See* Hessick, *supra* note 101, at 1516.

[118] *Id.* at 1517.

[119] *See supra* notes 96–99 and accompanying text.

sometimes correspond with a corpus analysis, while another individual's intuitions might differ.[120] The key question, though, is whether the corpus analysis is accurate in measuring collective usage. If so, any individual's differing intuitions do not matter. If not, corpus linguistics has no value and should not be used regardless of how notice and accountability concerns are resolved.

B. Arguments by Proponents that Exaggerate the Role of Corpus Linguistics

Corpus linguistics is thus an interpretive source (rather than an objective of interpretation) that provides evidence about collective usage relevant to the determination of ordinary meaning.[121] If a corpus analysis is used to determine the lexical meanings of statutory words and phrases, it should not displace any interpretive source other than dictionaries. Even then, considering the labor and technical knowledge required to produce a competent corpus analysis, it is unlikely that corpus analysis would completely displace dictionary definitions.[122] A structure of interpretation that would include corpus linguistics might therefore look like the following:

Statutory Interpretation—with corpus linguistics

Ordinary Meaning
—word meaning (dictionaries, intuition, precedent, dictionaries, *corpus linguistics*)
—sentential (and maybe broader) context
—textual canons
Communicative Meaning (language production and comprehension)
—legislative history
—textual canons
Legal Meaning
—absurdity doctrine
—substantive canons

[120] *See* Tobia, *supra* note 116 (describing how the results from corpus linguistics can differ from individual judgments about meaning).

[121] *See* Slocum & Gries, *supra* note 104, at 17.

[122] *See id.* at 15–16.

As the diagram indicates, the introduction of corpus linguistics should not automatically displace any other interpretive source. The interpretive sources relevant to *legal meaning*, such as substantive canons, are unaffected by corpus evidence, as are interpretive sources relevant to *communicative meaning*, such as legislative history. Interpretive sources relevant to ordinary meaning, such as textual canons, should similarly not be displaced because they address language systematicities not covered by a corpus analysis relating to a specific interpretive issue.[123]

1. Corpus Linguistics and Communicative Meaning

The above account, where corpus linguistics provides evidence about lexical meaning and does not displace other interpretive sources, is relatively modest. Some have argued for a much more expansive conception of corpus linguistics. Lee and Mouritsen, for instance, argue that (1) determining ordinary meaning is an empirical issue and is thus amenable to knowledge and processes from the field of linguistics;[124] (2) corpus linguistics is superior to existing methods of determining ordinary meaning;[125] and (3) the scope of potential application of corpus analysis is broad enough to help determine the intent of the legislature.[126]

One immediate problem with Lee and Mouritsen's view of corpus linguistics is the tension between their claim that ordinary meaning is an empirical issue and their focus on legislative intent. Significantly, they conflate ordinary meaning and communicative meaning, making the objective of meaning, in their view, the "'intended' meaning of the lawmaker," which, apparently, corresponds to the "communicative content" or "ordinary meaning" of statutory text.[127] While it is a mistake to conflate ordinary and communicative meaning,[128] under either concept statutory

---

[123] Of course, corpus linguistics can be used for purposes other than researching lexical meaning, such as determining whether a textual canon represents an accurate generalization about language usage. *See, e.g.*, Matthew J. Traxler et al., *Context Effects in Coercion: Evidence from Eye Movements*, 53 J. MEMORY & LANGUAGE 1, 2 (2005) (using corpus linguistics to determine that expressions with fully specified event structures are rare (i.e., are elided) "when the event is commonly associated with the noun").

[124] *See* Lee & Mouritsen, *supra* note 111, at 795.

[125] *See id.* at 794–95, 798 ("The problem is underscored by the tools (mis)used by judges to try to answer this empirical question . . . ."); *id.* at 867 ("The potential for subjectivity and arbitrariness is not heightened but reduced by the use of corpus linguistics.").

[126] *See id.* at 823–24, 853–56. There is unresolved tension between many of the article's bold premises about the value of corpus linguistics and its denouement that judges should consider corpus analysis as "something of a last resort" that is used only in a "relatively rare case." *Id.* at 872. One of the authors mentions that in his five years on the Utah Supreme Court, he has "employed such analysis only a very few times." *Id.* at 872 n.322.

[127] *Id.* at 792–94.

[128] *See supra* notes 34–38 and accompanying text.

interpretation may appear to be an empirical issue because the objective of interpretation is often framed as measuring the views of a certain community, such as a legislature. For Lee and Mouritsen, empirical means simply "the sense of a word or phrase that is most likely implicated in a given linguistic context."[129] Thus, using H. L. A. Hart's famous hypothetical, if a legal rule "forbids you to take a vehicle into the public park,"[130] an interpretive dispute about whether a "bicycle" is a "vehicle" may well seem like an empirical issue.[131]

The problem with the empirical view of interpretation is that a corpus analysis may be empirical in nature, but that does not mean a statutory interpretation is similarly empirical.[132] Statutory interpretation involves consideration of evidence of both general and specific language usage.[133] Corpus linguistics can provide important information about general language usage, but such evidence must be combined with consideration of the specific context of a statute. The latter inquiry is not determined through corpus analysis.[134] The empirical view thus fails to sufficiently account for judicial consideration of the specific context of a statute, especially if the goal is the ascertainment of communicative meaning.

In order for corpus linguistics to transform legal interpretation into an empirical issue, it would be necessary for a corpus analysis to displace all of the other interpretive sources that make statutory interpretation non-empirical. But corpus linguistics cannot provide various types of information that are crucial to legal interpretation. First, corpus linguistics cannot account for issues of *legal meaning*, which involve such things as substantive canons. For instance, a corpus analysis cannot determine whether the government's statutory interpretation has raised a serious constitutional issue.[135] Second, corpus analyses cannot displace various interpretive sources that determine *language production communicative meaning*.[136] For instance, legislative history allows the interpreter to consider

---

[129] *See* Lee & Mouritsen, *supra* note 111, at 795.

[130] *See* Hart, *supra* note 46.

[131] Hart viewed "bicycles" as within the "penumbra of debatable cases." *Id.*

[132] This is true in the same way that corpus linguistics is based on a theory of language but that does not make it a theory of legal interpretation.

[133] *See supra* notes 30–40 (describing the contextual nature of statutory interpretation).

[134] *See supra* notes 96–99 (describing the information provided by corpus linguistics).

[135] In fact, critics have argued that courts do a poor job of determining whether an interpretation raises a serious constitutional issue. *See, e.g.*, Neal Kumar Katyal & Thomas P. Schmidt, *Active Avoidance: The Modern Supreme Court and Legal Change*, 128 HARV. L. REV. 2109, 2122 (2015) ("The avoidance canon enables—even demands—sloppy and cursory constitutional reasoning.").

[136] *See supra* note 49 and accompanying text (describing *language production communicative meaning*).

the particularized context surrounding the enactment of a statute and make inferences about legislative intent.[137] This sort of information cannot be derived from corpus analysis.[138] Finally, even textual canons, many of which are arguably aspects of ordinary meaning, require a court to consider the particularized context of a statute, which is not the function of corpus analysis.[139] Thus, even determinants of ordinary meaning that are based on systematicities of language usage typically require courts to consider the context of the relevant statute.

Corpus linguistics therefore cannot displace all other interpretive sources, and may not displace any, which illustrates why corpus linguistics cannot transform statutory interpretation into an empirical issue.[140] Statutory-interpretation-is-empirical scholars might respond that other interpretive principles are not always applicable or that they change the "true meaning" of a provision.[141] Perhaps then corpus analysis can determine *communicative meaning* in at least some cases, which could at least give statutory interpretation some empirical basis. The flaw in such an argument is, again, the ineliminably contextual nature of interpretation and the need for any coherent interpretation to account for the particularized context of the relevant statute.[142]

---

[137] For an analysis of legislative history, see, e.g., James J. Brudney & Corey Ditslear, *The Decline and Fall of Legislative History? Patterns of Supreme Court Reliance in the Burger and Rehnquist Eras*, 89 JUDICATURE 220 (2006); Charles Tiefer, *The Reconceptualization of Legislative History in the Supreme Court*, 2000 WIS. L. REV. 205 (2000).

[138] That is, it cannot be derived from the kind of corpora that have thus far been constructed. *See* Lee & Mouritsen, *supra* note 111, at 828–35. Theoretically, it is possible to construct a corpus full of legislative history from various statutes. The relevance of such corpus analysis to the interpretation of a given statute would depend on a language production view of meaning, along with various fictional assumptions about the approval of the enacting Congress of legislative history unconnected to the provision at issue.

[139] *See* Anita S. Krishnakumar, *Backdoor Purposivism*, 69 DUKE L.J. 1275 (2020) (explaining that textual canons allow judges to use context to engage in purposive reasoning). Lee & Mouritsen make various arguments about the problems with textual canons, Thomas R. Lee & Stephen C. Mouritsen, *The Corpus and the Critics*, 88 U. CHI. L. REV. 275, 289–91 (2021), but such criticisms do not prove anything about the scope of corpus linguistics. Even if corpus linguistics were valid as applied to legal texts, and textual canons invalid, the corpus analyses performed by Lee & Mouritsen (and other legal scholars) do not produce the sorts of presumptions of language usage represented by textual canons, many of which, when applied, result in non-literal interpretations. *See* Tobia et al., *supra* note 81.

[140] That is, currently not all interpretive sources are empirically based, and an interpretive source would thus need to displace all non-empirically based interpretive sources in order to transform statutory interpretation into an empirical issue.

[141] Certainly, most (if not all) interpretive sources have a limited range of application, and one can always argue that the text has a 'correct' meaning that is independent of some interpretive source that reflects legal concerns. *See* Larry Alexander & Saikrishna Prakash, *"Is That English You're Speaking?" Why Intention Free Interpretation Is an Impossibility*, 41 SAN DIEGO L. REV. 967, 969 (2004) (arguing that the "actual meaning of a legal text—what its author(s) intended it to mean—might differ from the authoritative meaning that an authoritative interpreter gives it" (emphasis omitted)).

[142] The linguistic meaning of a legal text is not limited to the semantic meaning of the language but, rather, includes the pragmatic processes necessary to identify the meanings of the

Corpus analysis would need to take account of this particularized context for the corpus-linguistics-makes-statutory-interpretation-empirical argument to possess even a narrow plausibility. Corpus linguistics involves statistical and related analysis, however, not the qualitative examination of an individual statute.

2. Corpus Linguistics and Implied Exceptions to Statutes

There is thus a distinction between general and specific evidence of language usage that is crucial to statutory interpretation. But what if, corpus linguistics scholars might argue, a corpus contained scenarios that track quite closely the context of the relevant statute? Perhaps then a corpus analysis can reveal both general and specific information sufficient to determine the communicative meaning of the statute. While such information could be quite useful in certain scenarios, it would still not transform statutory interpretation into an empirical determination because it could not displace inferences based on consideration of the context of the particular statute at issue.

Lee and Mouritsen use Judge Richard Posner's "keep off the grass" hypothetical, as well as Hart's "no vehicles in the park" hypothetical, to argue that corpus linguistics can in fact determine the communicative meaning of a legal text.[143] The interpretive question in Posner's hypothetical is whether a sign in a park that says "keep off the grass" is properly interpreted to forbid such things as a grounds crew from cutting the grass, and in Hart's hypothetical Lee and Mouritsen focus on whether the prohibition applies to ambulances.[144] Lee and Mouritsen indicate that they understand an interpretive question "in light of its pragmatic context, which includes inferences about the place and manner of the utterance and presumed intentions of the speaker."[145] This understanding of meaning may be fine, but framing the interpretive process as including "inferences" about context and "presumed intentions" turns interpretation in a non-empirical direction. Furthermore, it is not clear how corpus analysis could assist an interpreter in making such inferences.

---

specific textual utterances of the legislature. While semantic meaning must in some ways account for context, identifying utterance meaning requires that particular consideration be made of context. *See* Scott Soames, *Deferentialism, Living Originalism, and the Constitution*, *in* THE NATURE OF LEGAL INTERPRETATION: WHAT JURISTS CAN LEARN ABOUT LEGAL INTERPRETATION FROM LINGUISTICS AND PHILOSOPHY 218–19 (Brian G. Slocum ed., 2017).

[143] *See* Lee & Mouritsen, *supra* note 111, at 824 (Posner's hypothetical); *id.* at 836 (Hart's hypothetical).

[144] *Id.* at 824, 836.

[145] *Id.* at 824.

Lee and Mouritsen nevertheless reason that it may be possible to examine the above interpretive questions "from a corpus-based perspective."[146] They reason as follows:

> If we had a large enough database, that contained a sufficient number of park prohibitions (together with references to groundskeepers, ambulances, etc.), we might be able to draw conclusions about the pragmatic circumstances in which such prohibitions are most commonly invoked and how they are most commonly interpreted. To find any ordinary exceptions to the "Keep off the grass" or the "no vehicles" rules we might look for park owners who have these rules in place. If park owners and municipalities routinely allow ambulances into their parks or routinely allow groundskeepers access, we can infer something about how these prohibitions are ordinarily used or understood. The point is that corpus analysis often contains at least some pragmatic data and is at least theoretically capable of providing information about the pragmatic context. But there is no guarantee that even a very large and targeted corpus would contain sufficient examples of circumstances with similar pragmatic content. And the question for corpus linguistics is how much of the relevant pragmatic context is reflected in the formal record found in the corpus.[147]

Despite the above arguments, if anything, the "keep off the grass" and "no vehicles" hypotheticals illustrate the inherent limitations of corpus analysis, as well as the nonempirical nature of statutory interpretation. Assuming, for a moment, that the sort of corpus that Lee and Mouritsen describe might exist, note that the interpretive questions relating to the two hypotheticals do not involve issues of explicit lexical meaning (i.e., the meaning of "grass" or "vehicle"). Rather, the focus is on intended exceptions to provisions that do not contain explicit exceptions. Consider the issues from a *reader comprehension communicative meaning* perspective. The availability of a rich trove of information about implied exceptions seems highly unlikely, but, presumably, the hope is that the corpus would contain sufficient reactions from ordinary citizens to the availability of implied exceptions so that one could get some sense of the *ordinary meaning* of the implied exceptions to such provisions.[148] If the corpus data does not involve public laws, however, there would be significant issues regarding whether ordinary exceptions to privately owned parks would be applicable to publicly owned parks.[149] Furthermore, even if available, the data would provide only generalized, indirect evidence about the implied exceptions, leaving the particularized

---

[146] *Id.* at 853.

[147] *Id.* at 853–54.

[148] Lee and Mouritsen concede this point implicitly by indicating that the corpus data would reveal how the hypotheticals "are most commonly interpreted." *Id.* at 853.

[149] It could well be that the contexts are so fundamentally different that few exceptions common to privately owned parks would be applicable to publicly owned parks.

context of the statute to be considered. An examination of that context may well reveal features of the overall statutory scheme that make the ordinary exceptions inapplicable. Corpus linguistics does not help, though, in analyzing this particularized context.[150]

Consider instead an approach from a *language production communicative meaning* perspective. Presumably, the research question would be whether the corpus contains sufficient indications of typical authorial intent regarding the availability of implied exceptions so that one could get some sense of the *ordinary meaning* of the implied exceptions to such provisions.[151] Even if obtainable, one may wonder how persuasive such information should be to a court's interpretation. As with *reader comprehension communicative meaning*, if the data from the corpus does not involve public laws, there would be significant issues concerning the inferences that can be made about legislative intent from nonlegislative scenarios. Even if we assume the corpus data involves public laws, reliance on information about typical legislative intentions would involve a likely legal fiction that the legislature is aware of the other provisions (which could be from other jurisdictions and involve statutes enacted long ago) and intends similar exceptions.[152] Furthermore, there may well be features of the statutory scheme that make the ordinary exceptions inapplicable. Again though, corpus linguistics does not help in analyzing the particularized context of the specific statute at issue.

In a broader sense, Lee and Mouritsen's hypothetical corpus full of contextually similar scenarios to actual legislation (sufficient to discern the scope of implied exceptions!) seems unlikely even when the statute at issue might conceivably contain the kind of language that would exist outside the law.[153] Statutory scenarios that have useful nonlegal parallels are even more unlikely. First, the "keep off the grass" and "no-vehicles-in-the-park" hypotheticals contain simple rules, but statutes are

---

[150] Corpus linguistics could, theoretically, create a presumption based on ordinary meaning (as happens with ordinary meaning generally), but the particularized context would always need to be examined in order to confirm or rebut the presumption.

[151] The corpus data might reveal, for example, that such provisions are normally intended to have certain exceptions or that, in any case, they are normally interpreted by relevant parties as having such exceptions.

[152] Consider that it is much more plausible to assume that a legislature is aware of and intends to enact the conventional meanings of words than it is to assume that a legislature is aware of the work of other legislatures and intends to enact those other legislative schemes into law. Thus, courts should be more willing to accept fictions about lexical meaning than about implied exceptions.

[153] For such a scenario to be plausible, the current drafting practices of legislatures would have to fundamentally change. *See generally* Peter M. Tiersma, *A Message in a Bottle: Text, Autonomy, and Statutory Interpretation*, 76 TUL. L. REV. 431 (2001) (describing the differences between legal and non-legal language).

typically written in language that is quite different from nonlegal language, making one-to-one comparisons difficult.[154] Statutory words can be given their ordinary meanings, of course, but replication of the exact language of a provision, let alone the entire context of a statute, is unlikely. Furthermore, many statutory schemes govern complex regulatory efforts, which do not have corollaries outside the law.[155] Certainly, a corpus could be constructed that would consist of statutes and other legal materials such as legislative history, but the corpus analysis would then involve a focus on language production with the attendant fictions of legislative intent already discussed.[156]

The use of corpus analysis does not therefore transform the determination of statutory meaning into an empirical issue. Rather, corpus linguistics is an interpretive source that provides information relevant to the ordinary meaning of statutory terms. Unlike various other interpretive sources such as legislative history, the main function of corpus analysis is to provide data about word meanings that cut across contexts.[157] Corpus linguistics thus can reveal important systematicities of language usage but is less useful to the determination of communicative meaning. Rather, any corpus analysis must be combined with an examination of the particularized context of a statute in order to determine the meaning of the relevant provision.

IV. SURVEY EVIDENCE WITHIN THE STRUCTURE OF INTERPRETATION

Corpus linguistics is thus like other traditional interpretive sources in the sense that it provides indirect evidence of a community's views regarding certain aspects of an interpretive dispute.[158] In contrast, a survey can be a direct way of measuring a particular group's views about a specific interpretive issue, as well as a way of providing information about a community's views about more general issues.[159] Increasingly, legal scholars have used survey evidence to provide evidence about both general issues and specific interpretive disputes. For instance, scholars have used survey evidence to assess the accuracy of corpus linguistics, as well as how ordinary people understand key phrases of statutory provisions,

---

154 *See id.*

155 *See* SLOCUM, *supra* note 28, at 239 (explaining that terms in legal texts often refer to intangible concepts that "do not exist outside of the law").

156 It may not after all provide an advantage to Westlaw.

157 *See supra* notes 96–99 and accompanying text.

158 *See supra* notes 61–89 and accompanying text.

159 *See generally* Kevin Tobia, *Experimental Jurisprudence*, 89 U. CHI. L. REV. (forthcoming 2022).

such as causal language like "because of" and "results from."[160] Most provocatively, some scholars have also proposed that surveys should be used to measure how ordinary people interpret and apply entire contracts and statutes to specific interpretive disputes.[161]

Unlike corpus linguistics, surveys can resolve the tension between empiricism and context by accounting for the particularized context of a statute.[162] A survey can show participants a statute (like Hart's no-vehicles-in-the-park hypothetical) and ask them to apply it to a specific situation (such as whether a "bicycle" is a "vehicle" or whether an "ambulance" is prohibited).[163] Along with the complete language of the provision, participants can be given the context of a statute.[164] Surveys thereby obviate the ordinary meaning standard because there is no need to make inferences from generalizations about language usage.[165] Instead, by asking participants to resolve specific interpretive disputes in light of the full context of the statute, survey evidence can measure the *reader comprehension communicative meaning* of a provision.[166] Thus, if the interpretive objective is the language comprehension of ordinary people, surveys can provide evidence that will measure that comprehension.

Even so, surveys likely cannot account for all of the traditional interpretive sources. Survey participants can be given the full context of the statute, as well as descriptions of all potentially applicable interpretive sources, but survey participants may not be competent to apply interpretive sources that measure language production rather than comprehension.[167] Considering some of these

---

[160] *See* Macleod, *supra* note 15, at 958–59, 1006–08 (analyzing how an ordinary reader would understand Title VII's language by asking ordinary readers to apply that language in context, drawing on a set of nationally representative survey experiments). *See generally* Tobia, *supra* note 116 (analyzing whether corpus linguistics provides accurate information about the meanings of words in legal texts).

[161] *See* Omri Ben-Shahar & Lior Jacob Strahilevitz, *Interpreting Contracts Via Surveys and Experiments*, 92 N.Y.U. L. REV. 1753, 1766–82 (2017); Shlomo Klapper et al., *Ordinary Meaning from Ordinary People*, U.C. IRVINE L. REV. (forthcoming 2021) (manuscript at 19–21) (using surveys to measure how ordinary people apply statutes to specific interpretive disputes) (on file with author).

[162] *See supra* Section III.B (describing why corpus analysis does not transform statutory interpretation into an empirical issue because it cannot account for the entire context of a statute).

[163] *See* Klapper et al., *supra* note 161, at 34.

[164] *See id.*

[165] *See supra* notes 28–36 and accompanying text (describing the ordinary meaning doctrine). The claim here is only that a contextually rich survey may eliminate the necessity of making inferences from general word meanings to the specific context of a statute. It may be that surveys require other generalizations, such as what participants do in a highly artificial situation (the survey) is a reflection of their actual understandings of a statutory term.

[166] *See supra* notes 44–48 and accompanying text.

[167] It is assumed, *arguendo*, that the full context of a statute can be provided to survey participants, but it is doubtful that a survey could supply the breadth of context that a judge may consider, which could include related and other provisions, extensive

interpretive sources underscores the extent to which legal training and knowledge is integral to statutory interpretation. For instance, a textual canon like *in pari materia* recognizes a presumption that the same words in related provisions should be given the same meaning.[168] Why should we assume that the views of survey participants regarding the importance of consistency in any given statutory scheme are more relevant or important than a judge's views, or even that such views would be of benefit to a judge?[169] Similarly, we would expect that judges are more competent in evaluating and understanding legislative history than ordinary survey participants.[170]

The same is true with respect to the interpretive sources that determine *legal meaning*. Survey participants are likely not capable of competently applying substantive canons like the avoidance canon.[171] Again, though, why would we assume that the views of survey participants regarding the application of substantive canons are more reliable than a judge's views, or even that such views would be of benefit to a judge?[172] Perhaps information about how ordinary people understand the law is useful to issues involving the legitimacy of interpretive rules. Thus, for instance, perhaps ordinary people expect that laws will operate only prospectively.[173] Such views may arguably serve to legitimize the presumption against retroactivity but would not mean that ordinary people possess the expertise required to ascertain whether a particular statutory application would operate retroactively.

Unless the objective of interpretation is narrowly defined *reader comprehension communicative meaning*, surveys therefore

---

legislative history of the current provision as well as other provisions, and precedent from various jurisdictions.

[168] *See supra* notes 82–84 and accompanying text.

[169] Consistency may not be important to ordinary people, but judges may understand the value of coherence and consistency in the law. In such situations, the jurisprudential question would be whether judges should feel constrained by the views of ordinary, nonlegal actors about issues outside their expertise.

[170] *See supra* notes 76–79 and accompanying text (explaining that it may be a legal fiction to assume that ordinary people would consult legislative history when interpreting a statute).

[171] *See supra* notes 51–52 and accompanying text.

[172] There are, however, applications of survey evidence that might be useful to judges. For instance, the canon of constitutional avoidance is controversial in part because critics claim that judges often disingenuously find clear provisions to be ambiguous, thereby authorizing application of the canon. *See* William K. Kelley, *Avoiding Constitutional Questions as a Three-Branch Problem*, 86 CORNELL L. REV. 831, 831–32 (2001) (calling for the abandonment of the avoidance canon in part because it "frequently results in questionable statutory interpretations"). Conceivably, survey participants could provide evidence about whether an ordinary person would find a particular provision to be ambiguous. Similarly, survey evidence may help judges evaluate whether a given textual canon is accurate. *See* Tobia et al., *supra* note 81.

[173] *See supra* notes 87–89 and accompanying text.

cannot in general transform an interpretive dispute into an empirical issue. Statutory interpretation is often a multilayered process that involves normative decisions, specialized legal competence, and inferences from context. Survey evidence, like corpus linguistics, raises important questions about the empiricism of statutory interpretation but, like in other areas of law, proponents of the empirical view may inappropriately view issues that are at least partly normative as ones that are entirely empirical.[174] Still, when interpreting statutes, judges often make assertions about an objectified person or community's views about meaning, and survey evidence can help evaluate those assertions. As such, they may play a valuable role in helping us evaluate the judiciary's rhetorical assertions about how ordinary people understand language or the law.

CONCLUSION

The possibility of data-driven interpretive sources will benefit statutory interpretation in various ways, including by encouraging scholars to think critically about the elements of interpretation. In particular, by providing empirical evidence of language usage, data-driven approaches may have a chastening effect on the current rhetoric-centric nature of interpretive argumentation. Yet, perhaps counterintuitively, data-driven approaches also illustrate the particularized nature of a statutory interpretation. Although data-driven approaches can make statutory interpretation more rigorous, the contribution cannot inform every level of the multilayered interpretive process. Normative principles, specialized legal competence, and inferences from context will continue to direct statutory interpretations.

---

[174] *See* Gregory C. Keating, *Reasonableness and Rationality in Negligence Theory*, 48 STAN. L. REV. 311, 337–39, 381 (1996) (arguing that the "average" in the average reasonable person doctrine is a normative one, established by an objective community standard that may or may not be representative of actual human actors).