

ARTICLES

INTERPRETING CONTRACTS VIA SURVEYS AND EXPERIMENTS

OMRI BEN-SHAHAR[†] & LIOR JACOB STRAHILEVITZ[‡]

Interpreting the language of contracts may be the most common and least satisfactory task courts perform in contract disputes. This Article proposes to take much of this task out of the hands of lawyers and judges, entrusting it instead to the public. The Article develops and tests a novel regime—the “survey interpretation method”—in which interpretation disputes are resolved through large surveys of representative respondents, by choosing the meaning that a majority supports. This Article demonstrates the rich potential for this method to examine variations of contractual language that could have made an intended meaning clearer. A similar survey regime has been applied successfully in trademark and unfair competition law for decades to interpret precontractual messages, and this Article shows how it could be extended to interpret contractual texts. The Article focuses on the interpretation of consumer contracts as the primary application of the proposed method, but demonstrates how the method could also apply to contracts between sophisticated parties. To demonstrate the technique, this Article applies the survey interpretation method to five real cases in which courts struggled to interpret contracts. It then provides normative, pragmatic, and doctrinal support for the proposed regime.

INTRODUCTION	1754
I. THE UNEASY PROCESS OF CONTRACT INTERPRETATION	1761
A. <i>Contract Interpretation and the Ambiguity Test</i>	1761
B. <i>The Challenges of Applying the Interpretation Doctrine</i>	1763
II. THE SURVEY INTERPRETATION METHOD	1766
A. <i>Trademark and Unfair Competition Surveys</i>	1769

[†] Leo and Eileen Herzel Professor of Law, University of Chicago.

[‡] Sidley Austin Professor of Law, University of Chicago. We would like to thank Ronen Avraham, Lisa Bernstein, Ruoying Chen, John Coyle, Meirav Furth, Mitu Gulati, Gillian Hadfield, Dave Hoffman, Ethan Lieb, Jennifer Nou, Richard McAdams, Ariel Porat, Geof Stone, George Triantis, Andrew Verstein, Dana Wagner, Tess Wilkinson-Ryan, Eyal Zamir, and participants in workshops and conferences at the University of Chicago, Duke University, Fordham University, Keio University (Tokyo), the University of Michigan, Oxford University, the 2017 Université de Lorraine and University of Chicago Law & Big Data Conference in Paris, the 2017 American Law & Economics Association annual meeting, and Sidley Austin LLP for helpful comments, Taylor Coles, Rafeh Qureshi, and Katie Ryan for outstanding research assistance, and the Coase-Sandor Institute for Law and Economics and Carl S. Lloyd Faculty Fund for research support. A special thanks is owed to Matthew Kugler for helpful assistance and advice with data collection and analysis. Copyright © 2017 by Omri Ben-Shahar & Lior Jacob Strahilevitz.

1.	<i>The Essential Role of Surveys</i>	1769
2.	<i>The Mechanics of Trademark and False Advertising Surveys</i>	1771
B.	<i>The Survey Method in Contract Interpretation</i>	1776
1.	<i>Whom to Survey?</i>	1776
2.	<i>What to Ask?</i>	1777
3.	<i>How to Determine a Winner?</i>	1779
4.	<i>Reliability of Survey Evidence</i>	1780
III.	THE SURVEY INTERPRETATION METHOD IN ACTION ...	1782
A.	<i>Methodology</i>	1783
B.	<i>Results</i>	1784
1.	<i>Experiment #1: Ambiguous Homeowner's Insurance</i>	1784
2.	<i>Experiment #2: Ambiguous Bonus Agreement</i> ..	1788
3.	<i>Experiment #3: Interpretation Against the Drafter</i>	1792
4.	<i>Experiment #4: A "Little Guy" Effect?</i>	1795
5.	<i>Experiment #5: Interpretation of Consumer Standard Form Contract</i>	1797
C.	<i>Summary</i>	1799
IV.	THE CASE FOR SURVEY INTERPRETATION	1802
A.	<i>Enforce the Terms People Expect</i>	1802
B.	<i>Costs of Interpretation</i>	1805
C.	<i>The End of "Normative" Interpretation</i>	1808
D.	<i>Simplifying the Contracts</i>	1810
E.	<i>Uncovering Sophisticated Meaning</i>	1813
V.	THE DOCTRINAL FOUNDATIONS OF THE SURVEY INTERPRETATION METHOD	1815
A.	<i>The Survey Method in Courts</i>	1815
B.	<i>Opting into the Survey Interpretation Method</i>	1820
1.	<i>Customizing Interpretation Law</i>	1820
2.	<i>Pretesting Contract Language</i>	1822
VI.	FURTHER EXTENSIONS AND CONCLUDING THOUGHTS ..	1824

INTRODUCTION

During the first half of the last century, plaintiffs seeking to prove trademark infringement knew what they had to do. The lynchpin of trademark litigation is consumer confusion, so a plaintiff seeking to establish such likelihood of confusion would call witnesses. Here is a consumer from Utah who will testify that the defendant's mark is confusingly similar to the plaintiff's mark, which she knows and trusts. Here is a witness from Florida who will attest to the same effect. I

have a shopper from Delaware who will describe buying the defendant's product by mistake, thinking it to be the plaintiff's product. And another from California. Witnesses would be marched in from all over. Everyone understood that the plaintiff (and defendant) are cherry-picking consumers to testify for or against confusion.

Then, in 1928, an enterprising lawyer in Delaware tried something novel. Instead of parading a couple dozen witnesses to testify, he brought in only one: an expert witness. The expert had surveyed hundreds of consumers about whether they were confused, producing an elegant study that generalized how average consumers responded to particular advertising messages and how many of them were confused by similar rival products. The court was not impressed, excluding the expert testimony as inadmissible hearsay. It insisted that it could only accept evidence from consumers testifying in open court.¹ Other courts followed suit.²

But in the decades that followed, federal courts began to reverse course.³ They came to recognize that expert witnesses of the sort called by the Delaware lawyer provided a much more reliable basis for determining whether two messages were likely to be confused with each other.⁴ They also came to realize that the fact that a handful of consumers would testify in court about having been deceived should not be dispositive.⁵ By the early 1960s, judicial opposition to consumer survey evidence had crumbled.⁶ Such evidence is now *de rigueur* in trademark and false advertising litigation.⁷

When it comes to contract interpretation, U.S. law may as well be stuck in 1928. The question of what a consumer contract means and

¹ *Elgin Nat. Watch Co. v. Elgin Clock Co.*, 26 F.2d 376, 376–78 (D. Del. 1928).

² *DuPont Cellophane Co. v. Waxed Prods. Co.*, 6 F. Supp. 859, 884–85 (E.D.N.Y. 1934), *modified on other grounds*, 85 F.2d 75, 80 (2d Cir. 1936).

³ *See, e.g., United States v. 88 Cases, More or Less, Containing Bireley's Orange Beverage*, 187 F.2d 967, 974 (3d Cir. 1951) (holding that a survey asking consumers what they believed the appellant's product contained did not constitute hearsay because it was not offered for the truth of the matter asserted, and allowing the survey to be admitted as evidence); *see also Schering Corp. v. Pfizer Inc.*, 189 F.3d 218, 224–25 (2d Cir. 1999) (summarizing the development of the case law).

⁴ *See Beverly W. Pattishall, Reaction Test Evidence in Trade Identity Cases*, 49 TRADEMARK REP. 145, 150–51, 150 n.22 (1959).

⁵ *Id.* at 148–49, 149 n.16.

⁶ *See Zippo Mfg. Co. v. Rogers Imps., Inc.*, 216 F. Supp. 670, 680–82 (S.D.N.Y. 1963) (stating that “[a]lthough courts were at first reluctant to accept survey evidence or to give it weight, the more recent trend is clearly contrary”).

⁷ *Schering Corp.*, 189 F.3d at 225 (acknowledging that surveys are “routinely admitted in trademark and false advertising cases” (citing *Bristol-Myers Squibb Co. v. McNeil-P.P.C., Inc.*, 973 F.2d 1033, 1043 (2d Cir. 1992); *PPX Enters., Inc. v. Audiofidelity Enters., Inc.*, 818 F.2d 266, 271 (2d Cir. 1987); *Nestle Co. v. Chester's Mkt., Inc.*, 571 F. Supp. 763, 769–70, 773–75 (D. Conn. 1983))).

what expectations it arouses among consumers is not determined based on how hundreds or thousands of surveyed representative consumers actually interpret the language at issue. Not even when there is a consensus among the respondents. Instead, contractual meaning is determined based on canons of interpretation, wispy policy arguments, judicial conjectures about what interpretations make business sense, or dictionaries. And when a court interprets a merchant contract based on the text's dominant usage in trade, the method for finding the trade usage is apt to be anachronistic. A handful of people involved in the industry or employed by the parties will be flown in to testify. Contract interpretation does not have to be that way. In this Article, we offer a better alternative.

Interpreting the language of contracts may be the most common and least satisfactory task courts perform in contract disputes. This Article proposes to take much of this task out of the hands of lawyers and judges, entrusting it instead to the relevant public. It develops and tests a novel regime—the “survey interpretation method”—in which interpretation disputes are resolved through surveys of representative populations, and ambiguity is measured or alleviated through the use of randomized experiments.

We approach this project with some sense of urgency. Contract interpretation is a body of law in dire need of new thinking. First, it is notoriously inconsistent. Different jurisdictions apply radically different approaches on the most fundamental question: What evidence should courts admit to inform their interpretation of contractual language? New York and California present two polar extremes, with the former admitting perhaps too little and the latter surely too much.⁸ And within state jurisdictions, courts are insubordinate. For example, the Illinois Supreme Court has instructed lower courts to consider only certain information, but to no avail.⁹

⁸ Compare *Mitchill v. Lath*, 160 N.E. 646, 549–50 (N.Y. 1928) (Lehman, J., dissenting) (arguing that the application of New York's parol evidence rule is too restrictive), with *Trident Ctr. v. Conn. Gen. Life Ins. Co.*, 847 F.2d 564, 569 (9th Cir. 1988) (arguing that California's erosion of the parol evidence rule is too permissive).

⁹ Compare *Air Safety, Inc. v. Teachers Realty Corp.*, 706 N.E.2d 882, 884 (Ill. 1999) (reaffirming a narrow parol evidence rule), with *Ortman v. Stanray Corp.*, 437 F.2d 231, 235 (7th Cir. 1971) (finding that even if the decision cannot be harmonized with “holdings of the various Illinois cases involving the parol evidence rule,” extrinsic evidence should always be admissible to determine the meaning of an integrated agreement). A practitioner's treatise concludes that “despite the Illinois Supreme Court's consistent reaffirmation of the four-corners test . . . , the Illinois appellate courts remain split on the use of those approaches.” Christofer R. Dunsing & Thomas R. Stilp, *The Parol Evidence Rule and Contract Interpretation*, in *CONTRACT LAW* § 11.15 (Karen F. Botterud ed., 2016 ed.).

Second, contract interpretation law is in trouble because it is overly complex. Courts employ a mishmash of conflicting methodologies—including textualism, formalism, purposivism, and functionalism—to elicit the meaning of texts.¹⁰ They rely on various policy goals—accuracy, reduction of transaction and litigation costs, improved drafting, and rewarding information exchange—to expand or shrink the scope of their contextual inquiry.¹¹ And they fall back on archaic canons of interpretation that have essentially nothing to do with the way people read prose.¹²

Third, and largely due to its inconsistency and complexity, contract interpretation is unpredictable. Because the amount of evidence and the nature of justifications that courts marshal to interpret the language is so temperamental, it is hard for contracting parties to anticipate the litigated outcome of their contracts. Contracting in the shadow of this interpretive risk becomes needlessly costly. In their (sometimes futile) effort to reduce this risk, drafters of contracts are writing longer and longer documents, further divorcing legal and lay meanings.¹³ It is thus not surprising that, reading a lengthy boilerplate, only nine percent of consumers understood that their contract had a legally enforceable mandatory arbitration clause.¹⁴

Interpretation risk is not the only cost imposed on parties by the present state of contract interpretation doctrine. Litigation itself has become expensive, with parties sometimes spending many years and

¹⁰ See Alexander Volokh, *Choosing Interpretive Methods: A Positive Theory of Judges and Everyone Else*, 83 N.Y.U. L. REV. 769, 779–80 (2008) (giving examples of the same provision being interpreted differently under the influence of different interpretive methods).

¹¹ See generally E. ALLAN FARNSWORTH, *CONTRACTS* 439–41 (4th ed. 2004); Keith A. Rowley, *Contract Construction and Interpretation: From the “Four Corners” to Parol Evidence (and Everything in Between)*, 69 MISS. L.J. 73 (1999) (discussing a number of rules of contract interpretation).

¹² See, e.g., *Rodriguez v. Miranda*, 507 S.E.2d 789, 793 (Ga. Ct. App. 1998) (describing a Georgia rule that any ambiguity should be interpreted against foreclosures); *Rogers v. Allied Mut. Ins. Co.*, 520 N.W.2d 614, 617 (S.D. 1994) (discussing the rule of the last antecedent used to interpret modifier words in contracts). See generally FARNSWORTH, *supra* note 11, at 456–61 (explaining the major rules used by judges to interpret contractual language); Edwin W. Patterson, *The Interpretation and Construction of Contracts*, 64 COLUM. L. REV. 833 (1964) (discussing a number of rules of contractual interpretation).

¹³ Todd D. Rakoff, *Contracts of Adhesion: An Essay in Reconstruction*, 96 HARV. L. REV. 1173, 1226 (1983) (predicting that “the standard [contract] forms—because they are drafted to cover many such contingencies—are likely to be long and complex”). See generally, OMRI BEN-SHAHAR & CARL E. SCHNEIDER, *MORE THAN YOU WANTED TO KNOW: THE FAILURE OF MANDATED DISCLOSURE* 18–23 (2014) (describing the length of standard form contracts).

¹⁴ Jeff Sovern et al., “*Whimsy Little Contracts*” with *Unexpected Consequences: An Empirical Analysis of Consumer Understanding of Arbitration Agreements*, 75 MD. L. REV. 1, 47 (2015).

millions of dollars fighting—not to persuade juries, but to convince judges what evidence to admit and what procedures to follow in interpreting the contract.¹⁵ And these problems have only been exacerbated in the digital era as contracts have grown longer.

The unhappy state of interpretation doctrine in contract law does not diminish its practical importance. As mass-market contracts are growing longer and more numerous, they also boldly override entire codes of default rules.¹⁶ Contracts are assuming a more prominent role in regulating market transactions,¹⁷ the transactions are becoming increasingly complicated, and their interpretation is becoming a battle of greater stakes.

This paradoxical combination—an increasingly archaic doctrine called upon to resolve increasingly important issues—suggests that it is perhaps time for a major new move. This Article proposes such a move: *outsourcing* the interpretation task by handing it over to a simpler, more predictable, and arguably cheaper process. Instead of asking judges and juries to interpret contracts, the meaning of disputed contractual clauses should be determined by polling a large representative sample of disinterested respondents. Let majorities of survey respondents decide. For consumer contracts (our primary focus in this Article), that entails polling a representative sample of consumers. For contracts involving sophisticated parties, we propose surveying dozens of disinterested participants in the relevant industry when feasible.

The survey interpretation method might appear to be a radical departure from the procedures governing existing contract litigation. In presenting this method as a practical alternative to court- and canon-based interpretation, we ground it on several supporting pillars. As our prologue revealed, a very similar method is widely used with satisfactory results in resolving interpretation disputes over the meaning of advertising messages and descriptions of products—texts closely related to contracts.¹⁸ The opposing parties in such cases (pri-

¹⁵ See, e.g., *TruServ Corp. v. Morgan's Tool & Supply Co.*, 39 A.3d 253, 256, 265 (Pa. 2012) (reversing and remanding for new proceedings in a breach of contract action that had been pending for twelve and a half years); *S'holder Representative Servs., LLC v. Airbus Ams., Inc.*, 791 S.E.2d 724, 727 (Va. 2016) (involving the breach of a merger agreement in which the court awarded the prevailing party attorney's fees of nearly \$3.9 million).

¹⁶ MARGARET JANE RADIN, *BOILERPLATE: THE FINE PRINT, VANISHING RIGHTS, AND THE RULE OF LAW* 182–85 (2013) (explaining how mass-market contracts diverge from default contract rules by introducing provisions such as choice of forum clauses, warranty disclaimers, and others).

¹⁷ Omri Ben-Shahar, *Regulation Through Boilerplate: An Apologia*, 112 MICH. L. REV. 883 (2014).

¹⁸ See *Schering Corp. v. Pfizer Inc.*, 189 F.3d 218, 225 (2d Cir. 1999).

marily in trademark and unfair competition disputes) are expected to produce surveys about the meaning that consumers assign to these messages and the likelihood that consumers will be confused by various aspects of these communications. When such surveys are done well, courts devotedly rely on them. We argue that if the survey method is deemed reliable in interpreting *precontractual* statements and communications, then they ought to be equally attractive in interpreting *contractual* communications—namely, the promises made in the contract.

While the use of consumer surveys in trademark law provides important cross-substantive inspiration for our proposal, we rely on three other types of support to make the case for the survey interpretation method: pragmatic, normative, and doctrinal. The core pragmatic argument is that interpretation surveys are practical and reliable. While we defer some of the nitty-gritty survey design questions, we nevertheless tackle several key methodological issues. We bolster the pragmatic argument by running a pilot study applying the survey method to resolve actual interpretation disputes. Specifically, we picked several real cases in which courts struggled to interpret contracts coherently, and conducted large-scale surveys asking lay respondents to interpret the same language. We were able to compare the results of the traditional and the survey methodologies, and highlight the attractive features of the latter in this Article. We also used the survey method to test whether refinements in the contract language shown to respondents elicit different majority interpretations. We did this to identify potential biases of respondents, and also to compare the contractual text with variations that might improve its clarity. Such results—and especially the proven existence of (unused) clearer language—provide a more rigorous and first-of-its-kind foundation for “whose meaning prevails” doctrines like *contra proferentem* and mutual mistake.

The second form of support is normative. The survey method advances a particular conception of meaning: attaching to contracts the understanding assigned by those for whom they are written. Surveys that poll a sample of the intended audience capture that meaning more accurately than a judge’s imagination. This goal is achieved when contract language aimed at laypeople (like consumers and most employees) is interpreted by surveys of the general population, and contracts aimed at particular sectors (like specialists, merchants, and investors) are interpreted by surveys of sector members. In so doing, our methodology takes seriously the black letter doctrine that contract formation requires an objective “meeting of the minds,” and uses hard data as evidence of what the parties to the con-

tract actually expected (or would have expected had they read the text).¹⁹ By creating a reliable mechanism to evaluate the objective meaning, courts can be more certain that the terms they are enforcing are the ones to which both parties actually would have assented.

Interpretation via surveys would reduce the opportunity for courts to engage in “normative interpretation”—choosing interpretations that advance other goals. It would also promote important qualities. One such quality is the incentive to draft short, simple, and widely understandable contractual text *ex ante* in order to reduce the risk of misinterpretation. Another quality is low litigation cost. In the era of online panels, the cost of resolving interpretation disputes through surveys is potentially far lower than the cost of the alternative—with lawyers racking up billable hours to canvass precedents, canons, and context to sway trial courts. Finally, survey-interpreted consumer contracts are predictable, because they can be pretested. This also allows for more flexibility because contracting parties no longer have to stick with old, previously interpreted boilerplates to secure a known meaning.

The third and last defense of the survey interpretation method is doctrinal. Here, we make two separate arguments. One argument is conceptual. While the survey interpretation method is a procedural innovation, we argue that it is substantively consistent with some existing law. We show that the method shares the premises and foundations of longstanding contract and evidence doctrines.

Another doctrinal argument is founded on the power of contracting parties to customize litigation procedures. If the existing law of contract interpretation is just a default rule, parties could opt out of court-based interpretation and opt into the survey interpretation method. Not just could; they should (recall the aforementioned list of desirable effects). To reduce *ex post* battles of surveys, the parties could specify an exclusive survey procedure they expect the court to recognize and may even select the survey company in advance. This would also enable the parties to put the language of the contract to the test, *ex ante*. Businesses already field test every aspect of their products before releasing them to the public. It would be straightforward to field test their consumer contract language as well.

In developing the case for the survey interpretation method, this Article is structured as follows. Part I states the problem: Existing contract interpretation law is inconsistent and complex, and it gener-

¹⁹ See Margaret Jane Radin, *Boilerplate Today: The Rise of Modularity and the Waning of Consent*, 104 MICH. L. REV. 1223, 1231–32 (2006) (bemoaning the trend of treating the meeting of the minds as a legal fiction).

ates too much litigation at too high a cost. Part II introduces the solution, the survey interpretation method, and shows how a similar approach has worked in trademark litigation. Parts III, IV, and V present the three types of supporting arguments for the proposed method—pragmatic, normative, and doctrinal. Finally, Part VI concludes with further extensions and remarks.

I

THE UNEASY PROCESS OF CONTRACT INTERPRETATION

[I]nterpretation of the contract is no formal or mechanical task. On the contrary, it is one of the most intractable tasks which a court has to face, and it is not made easier by the inadequacy of the rules which the courts have forged to assist them.

— P.S. Atiyah²⁰

We begin by identifying the problem: Existing interpretation doctrines are difficult to apply and lead to costly litigation with unpredictable outcomes. To demonstrate this we focus on the leading exemplar of the existing interpretation method: the plain meaning rule. This is the idea that contractual text has to be interpreted according to its most straightforward meaning. Our interest is not so much in the substance of the plain meaning test, but in the process of its implementation—how courts go about figuring out the plain meaning of contractual terms. We show that the process is plagued by methodological complexity and murkiness. This brief discussion will set the stage for the main contribution of this Article, which begins in Part II, exploring a novel substitute for the prevailing approach.

A. *Contract Interpretation and the Ambiguity Test*

The basic question a court asks in interpreting the meaning of contractual language is how people entering such a contract understand it—what expectations this language evokes among the parties.²¹ “The cardinal rule in the interpretation of contracts is to ascertain the intention of the parties. To this rule all others are subordinate.”²²

Words, though, rarely have singular meaning. No matter how carefully drafted, the reasonable understanding of a word or phrase often depends on the circumstances in which it was used when the contract was made. A buyer and seller may agree on the sale of “Grade A chicken,” but does that mean only fresh young broilers, or

²⁰ P.S. ATIYAH, AN INTRODUCTION TO THE LAW OF CONTRACT 195 (5th ed. 1995).

²¹ FARNSWORTH, *supra* note 11, at 445–52.

²² *City of Grosse Pointe Park v. Mich. Mun. Liab. & Prop. Pool*, 702 N.W.2d 106, 113 (Mich. 2005) (quoting *McIntosh v. Groomes*, 198 N.W. 954 (Mich. 1924)).

also stewing chicken?²³ An airplane insurance contract may stipulate that it covers only accidents occurring “within the United States and its territories,” but does this exclude a plane crashing in international waters while en route from Miami to Puerto Rico?²⁴

Contractual disputes regularly arise over the meaning of a provision that is capable of being reasonably understood in more than one way. A judge often has to decide, as a matter of law (reviewable by an appellate court), whether the term is indeed susceptible to such dispute over meaning. If it is, its interpretation becomes a question of fact to be resolved via trial, where the multitude of surrounding circumstances including precontractual communications, drafting history, evidence about goals the parties had under the contract, past and concurrent dealings, and trade norms may be presented.

The essence of the plain meaning rule is in guiding the judge making this initial determination—whether a term is ambiguous. Should the judge go beyond the dictionary meaning of contract text and examine, in addition, the context in which the language was used to determine whether more than one meaning is plausible? How much of the potential evidentiary thicket may the judge invoke, or even consider, in making this determination?

Courts are hopelessly split along the continuum between two polar approaches to this fundamental dilemma. At one end is the narrow “textual” approach of traditional common law—sometimes known as the “four corners test”²⁵—requiring the judge to ignore extrinsic evidence and refer only to the text in deciding whether ambiguity exists. This formalistic approach presumably reduces litigation, but it also causes courts to sometimes enforce bargains that the parties never intended. At the other end is the more expansive “contextual” approach of the Second Restatement and the Uniform Commercial Code, instructing the judge to consider interpretations based on at least some surrounding circumstances for the purpose of determining whether the term is ambiguous.²⁶ This fact-intensive approach

²³ *Frigalment Importing Co. v. B.N.S. Int'l Sales Corp.*, 190 F. Supp. 116, 117 (S.D.N.Y. 1960).

²⁴ *Vargas v. Ins. Co. of N. Am.*, 651 F.2d 838, 839 (2d Cir. 1981).

²⁵ *Saddler v. Nat'l Bank of Bloomington*, 85 N.E.2d 733, 740 (Ill. 1949); *Stewart v. McChesney*, 444 A.2d 659, 661 (Pa. 1982); *Treemont, Inc. v. Hawley*, 886 P.2d 589, 592–93 (Wyo. 1994).

²⁶ RESTATEMENT (SECOND) OF CONTRACTS §§ 212–216 (AM. LAW INST. 1981); U.C.C. § 2-202 (AM. LAW INST. & UNIF. LAW COMM'N 1977); *see, e.g.*, *Pac. Gas & Elec. Co. v. G.W. Thomas Drayage & Rigging Co.*, 442 P.2d 641, 644 (Cal. 1968) (holding that the test whether extrinsic evidence should be admissible is whether the evidence is “relevant to prove a meaning to which the language of the instrument is reasonably susceptible”).

improves accuracy, but at the cost of lengthier and more expensive litigation.

The same question—what information to rely on when determining whether the meaning of a term is ambiguous—arises in the application of other interpretation doctrines as well. One example is the parol evidence rule, which prevents a party from introducing extrinsic evidence of prior understandings when a written contract integrates the entire agreement.²⁷ As with the plain meaning rule, the judge has to decide whether to admit extrinsic evidence.²⁸ Likewise, the doctrine of reasonable expectations allows the court to bridge the perfunctory letter of the contract and its spirit by enforcing a meaning consistent with the intentions of the parties as evidenced by surrounding circumstances.²⁹ In most states, courts apply such expansive interpretation only after determining that the text itself is ambiguous and susceptible to this meaning.³⁰ These courts face the same methodological challenge: Is a term ambiguous? How much extrinsic evidence should be employed in determining whether it is?

B. The Challenges of Applying the Interpretation Doctrine

The legal framework that governs the interpretive task is concentrated around the *ambiguity* test. The court's role is to determine whether a term is sufficiently clear and has, as a matter of law, only one reasonable meaning.

The ambiguity test is perhaps the most practically important tool that courts adjudicating contracts disputes are asked to use, but it is also the least theoretically satisfying. This mismatch between the test's widespread usage and its impoverished theoretical underpinnings are in part due to the idiosyncratic nature of the interpretation enterprise. Determining whether a term is ambiguous is a task for which prior resolution of earlier cases provides limited, if any, guidance. Even if the same language was used previously, it was wrapped within different overall text, used among different parties at a different time, and illuminated by different (alleged) contextual clues. Interpretation is a task that requires judges to apply intuition and common sense,

²⁷ RESTATEMENT (SECOND) OF CONTRACTS § 213 (AM. LAW INST. 1981).

²⁸ FARNSWORTH, *supra* note 11, at 418–22.

²⁹ RESTATEMENT (SECOND) OF CONTRACTS § 211(3) (AM. LAW INST. 1981); *see, e.g.*, C & J Fertilizer, Inc. v. Allied Mut. Ins. Co., 227 N.W.2d 169, 176 (Iowa 1975) (rejecting the plain meaning rule in favor of the reasonable expectations test).

³⁰ Roger C. Henderson, *The Doctrine of Reasonable Expectations in Insurance Law After Two Decades*, 51 OHIO ST. L.J. 823, 827–28 (1990). *See also* RESTATEMENT OF THE LAW OF LIABILITY INSURANCE § 3(2) (AM. LAW INST., Proposed Final Draft 2017) (providing that insurance policies are to be interpreted according to their plain meaning).

namely mental processes that vary with the judges' backgrounds and experiences.³¹

The ambiguity test leads to ad hoc and unsatisfactory results for reasons that extend beyond the idiosyncrasies of cases or judges. The most perplexing challenge under this test is to define boundaries of permissible extrinsic evidence. How much information on surrounding circumstances ought to be utilized in determining whether the text is ambiguous? The extreme positions—"no extrinsic information" or "all extrinsic information"—are neither possible nor desirable. Even rigid textual approaches invoke some assumptions and experience that are not explicitly stated in the text—it is impossible and often silly to understand language "in a vacuum,"³² without a contextual baseline. For example, some extrinsic information is intuitively invoked to know that "Grade A chicken" refers to supply chain standards and not to the hygiene score of a Chick-fil-A outlet. And conversely, the entire distinction between questions of law and those of fact would collapse, and the reliability of written contracts would subside, if written terms were never litigation-proof and if the court, at the pre-trial stage, was required to examine all extrinsic evidence.

In theory, in order to determine how much information to admit in resolving the ambiguity test, the court can balance the cost and benefit of the marginal bit of evidence. Additional extrinsic data should be allowed only if the cost of evaluating it (lengthier proceedings) is less than the expected benefit (greater accuracy). But courts do not have the tools to perform this tradeoff, particularly because they are ill-equipped to ascertain the benefit of information prior to its acquisition. A party may argue that the meaning of "Grade A chicken" can only be established by evidence of trade usage, but it is not clear until such evidence is brought whether it would indeed shed light on the meaning. The value of information is hard to assess until it is acquired, by which time it is too late to conclude that it is not worth the cost.

This absence of clear criteria—which information to admit to resolve the ambiguity test—has led to several problematic features of existing law. First, courts are hopelessly split, even within jurisdictions, about the proper role of extrinsic evidence and the proper scope of the plain meaning rule. For example, while the Illinois Supreme Court has repeatedly reaffirmed the strict approach of the four corners

³¹ See 3 ARTHUR LINTON CORBIN, CORBIN ON CONTRACTS § 535 (rev. ed. 1960) ("It is true that when a judge reads the words of a contract he may jump to the instant and confident opinion that they have but one reasonable meaning and that he knows what it is.").

³² *Ortman v. Stanray Corp.*, 437 F.2d 231, 234 (7th Cir. 1971).

test,³³ lower courts often follow a more lenient and contextual approach, with occasional approval from the Seventh Circuit.³⁴ One cannot shake the cynical impression that the choice of approach—including more versus less extrinsic information to determine whether the language is susceptible to more than one meaning—is merely a label placed on the *conclusion* of the interpretive inquiry rather than a functional test to help resolve it.

The murkiness of the information test forces courts ill-equipped to make the cost/benefit tradeoffs to resort to arbitrary categorical classifications. For example, many courts distinguish between evidence of “surrounding circumstances” (permitted) versus evidence of “prior negotiations” (prohibited).³⁵ These classifications lead courts to make contradictory statements, sometimes in the same decision—like the following language from New York’s highest court: Ambiguity “is determined by looking within the four corners of the document, not to outside sources,” but in “deciding whether an agreement is ambiguous courts ‘should . . . consider the relation of the parties and the circumstances under which it was executed.’”³⁶

The difficulty in applying the ambiguity test is rooted in a fundamental and unresolved distinction that underlies the entire interpretive methodology—is interpretation a question of law or a question of fact? Any attempt to divide the labor between law/fact and judge/jury is constricting the process with an all-or-nothing choice that poorly fits the underlying challenges—namely, what intermediate amount of information is optimal in interpreting the contract, and what expertise is needed to evaluate this information? The confusion over the law/fact boundary runs deep in contract law. The Restatement of Contracts endorses an intermediate solution—while it classifies interpretation as a “question of law,” it also stipulates that it ought to be

³³ See, e.g., *Gallagher v. Lenart*, 874 N.E.2d 43, 63 (Ill. 2007) (affirming appellate court decision fully giving effect to straightforward language in a workers’ compensation settlement contract); *Air Safety, Inc. v. Teachers Realty Corp.*, 706 N.E.2d 882, 884 (Ill. 1999) (holding that a court must look at the language of the contract alone).

³⁴ See *Dunsing & Stilp*, *supra* note 9, §§ 11.5, 11.17 (describing the approaches taken by Illinois lower courts and the Seventh Circuit to the admission of extrinsic evidence).

³⁵ FARNSWORTH, *supra* note 11, at 464; see also *Fla. E. Coast Ry. Co. v. CSX Transp.*, 42 F.3d 1125, 1129 (7th Cir. 1994); *W.W.W. Assocs. v. Giancontieri*, 566 N.E.2d 639, 642 (N.Y. 1990); *Steuart v. McChesney*, 444 A.2d 659, 663 (Pa. 1982); *Berg v. Hudesman*, 801 P.2d 222, 230 (Wash. 1990); David E. Pierce, *Defining the Role of Industry Custom and Usage in Oil and Gas Litigation*, 57 S.M.U. L. REV. 387, 398 (2004) (stating that “many courts would exclude prior negotiation evidence under the plain meaning rule” (internal quotations omitted)).

³⁶ *Kass v. Kass*, 696 N.E.2d 174, 180 (N.Y. 1995); see also *Giancontieri*, 566 N.E.2d at 642 (“Evidence outside the four corners of the document as to what was really intended but unstated or misstated is generally inadmissible to add to or vary the writing.”).

determined by the “trier of fact” when the meaning depends on the credibility of, or the choice among reasonable inferences from, extrinsic evidence.³⁷ But courts are far from uniform in their treatment of this distinction.³⁸

All this uncertainty increases litigation costs. In requiring the judge to make an upfront determination about whether a contractual term is ambiguous, the present methodology induces the parties to spend significant resources to sway the judge that their preferred interpretation is at least plausible to move beyond summary proceedings. Contract disputes are costly, and these costs distort the outcomes by pressuring a party to settle in order to avoid lengthy and unpredictable litigation. These costs also turn the rationale of the plain meaning rule on its head. Recall that the primary justification for a textual approach is the saving of litigation costs involved in reviewing all the extrinsic evidence. But the battle over the ambiguity test could cost just as much. A plain meaning rule applied strictly might reduce litigation costs. Yet it does not appear that any U.S. jurisdiction has a sound basis for characterizing its contract law as applying the plain meaning rule in that way.

II

THE SURVEY INTERPRETATION METHOD

In Part I we described the problem. We now present the solution: a novel approach to the interpretation of contracts. To determine the meaning of a text, courts ought to rely on surveys. A survey would ask a pool of respondents, who resemble the contracting parties, what meaning they assign to the language. If the respondents are about evenly split, the term should be regarded as ambiguous. But if one meaning garners noticeably greater support among the survey respondents, it presumptively ought to prevail.

There are flickers of receptivity to this idea in contract doctrine. We discuss them in Part V. A couple of law review articles have flagged the idea of a more empirically inclined approach to contract interpretation.³⁹ And as then-Judge Sotomayor has noted, while the

³⁷ RESTATEMENT (SECOND) OF CONTRACTS § 212(2) & cmt. d. (AM. LAW INST. 1981).

³⁸ FARNSWORTH, *supra* note 11, at 477–78 (“The traditional view is that interpretation is generally for the judge. . . . Some courts, however, have shown greater willingness to send questions of interpretation to the jury.”).

³⁹ The paper that comes closest to advancing the argument we make here is Michelle Boardman, *Insuring Understanding: The Tested Language Defense*, 95 IOWA L. REV. 1075 (2010). Boardman’s argument is limited to the insurance contract context, but she was apparently the first scholar to argue that when an insurance company has pre-tested insurance policy boilerplate by having disinterested consumers read it, the company could invoke data about how respondents understood the language as a defense in subsequent

use of surveys is most widespread in trademark litigation, surveys have been admitted for other limited purposes in cases involving the Fair Housing Act, the Equal Protection Clause, and the Equal Educational Opportunities Act.⁴⁰ But the primary inspiration for the use of the survey method to interpret contracts comes from trademark and unfair competition law.

Contractual terms are one way in which parties communicate their promises. Another channel of communication is precontractual “marketing”—the representations that a party makes to its potential contractual counterparts through advertising, branding, and various statements and disclosures. In adjudicating the meaning of these communications and the expectations they convey to their recipients, courts have long relied on surveys.⁴¹ In a typical trademark infringe-

litigation alleging that the policy was ambiguous. *Id.* at 1099–103; *see also* Eric A. Zacks, *The Restatement (Second) of Contracts § 211: Unfulfilled Expectations and the Future of Modern Standardized Consumer Contracts*, 7 WM. & MARY BUS. L. REV. 733, 792 (2016) (noting the appeal of empirical approaches to contract interpretation). Another strand of the literature advocated the use of consumer surveys in the design of contract disclosures and default rules. *See, e.g.*, Ian Ayres & Alan Schwartz, *The No-Reading Problem in Consumer Contract Law*, 66 STAN. L. REV. 545 (2014) (advocating empirical testing to identify surprising and problematic provisions in standard form contracts, against which consumers ought to be warned); Ariel Porat & Lior Jacob Strahilevitz, *Personalizing Default Rules and Disclosure with Big Data*, 112 MICH. L. REV. 1417, 1419–20 (2014) (advocating the use of surveys to identify the majoritarian preferences for the design of granular default rules). Finally, survey methods have been used to measure consumers’ awareness of various contractual features. *See, e.g.*, Ann Morales Olazabal et al., *Frequent Flyer Programs: Empirically Assessing Consumers’ Reasonable Expectations*, 51 AM. BUS. L.J. 175, 248 (2014) (demonstrating a contrast between consumer expectations under frequent flyer programs and the actual contract terms); Lawrence Solan et al., *False Consensus Bias in Contract Interpretation*, 108 COLUM. L. REV. 1268 (2008) (comparing the presence of decisional biases among lay respondents and judges). A survey method of interpretation was also proposed recently in a German article, Alexander Stöhr, *Determining Transparency Under Section 307(1)(2) Civil Code: The Case for an Empirical Approach*, 216 ARCHIV FÜR DIE CIVILISTISCHE PRAXIS 558 (2016) (in German). *But see* Hanjo Hamann & Leonard Hoeft, *An Empirical Approach to Civil Law: Adding Realism and Methodological Rigor to Legal Analysis?*, 217 ARCHIV FÜR DIE CIVILISTISCHE PRAXIS (forthcoming 2017) (in German).

⁴⁰ *Schering Corp. v. Pfizer Inc.*, 189 F.3d 218, 225 (2d Cir. 1999) (Sotomayor, J.) (first citing *Keith v. Volpe*, 858 F.2d 467, 479–81 (9th Cir. 1988); and then citing *Debra P. v. Turlington*, 730 F.2d 1405, 1412–14 (11th Cir. 1984)).

⁴¹ Courts admit surveys as evidence under the state-of-mind exception to the hearsay rule, Federal Rule of Evidence 803(3). *See* Todd Anten, *In Defense of Trademark Dilution Surveys: A Post-Mosley Proposal*, 39 COLUM. J.L. & SOC. PROBS. 1, 2 (2005) (“Consumer surveys have always played prominent roles in trademark disputes, helping judges peek into the minds of consumers.” (citing Neal Miller, *Facts, Expert Facts, and Statistics: Descriptive and Experimental Research Methods in Litigation*, 40 RUTGERS L. REV. 101, 137 (1987))); *see also* Robert C. Bird & Joel H. Steckel, *The Role of Consumer Surveys in Trademark Infringement: Empirical Evidence from the Federal Courts*, 14 U. PA. J. BUS. L. 1013, 1025 (2012) (“Courts have called consumer surveys some of the most direct and persuasive evidence available to establish trademark infringement.” (first citing

ment or false advertising lawsuit, where the court has to interpret the meaning of a public communication, the outcome turns on the results of a survey that asks consumers what the message meant to them.⁴² In fact, surveys are such an essential type of evidence that failure to conduct them or to demonstrate their favorable findings may be fatal to a claimant's case.⁴³ Their use in competitors' disputes to decipher how people interpret communications is de facto "black letter law."⁴⁴

The gist of the approach presented in this Part is to extend the existing survey methodology, currently utilized to interpret important precontractual communications, into the interpretation of contract language. Contract promises, like marketing communications, are messages transmitted by one party to another, which create contractual obligations. When a business drafts a promise (or a disclaimer) into the standard contract terms, it is communicating with the same customers to whom advertising and product claims are made. If surveys are used to interpret what businesses *claim*, why not also what they *disclaim*? More generally, if the survey method is a reliable method to elicit the impact of the informal marketing-phase messages on the audience, why isn't it equally instructive for deciphering the meaning of drafted contractual terms? We therefore begin in Section II.A by reviewing the survey method in trademark and unfair competition law. Section II.B then presents the survey method in contracts disputes.

Checkpoint Sys., Inc. v. Check Point Software Techs., Inc., 269 F.3d 270, 283 n.10 (3d Cir. 2001); then citing *Vision Sports, Inc. v. Melville Corp.*, 888 F.2d 609, 615 (9th Cir. 1989); and then citing *McNeil Nutritionals, L.L.C. v. Heartland Sweeteners L.L.C.*, 566 F. Supp. 2d 378, 392 (E.D. Pa. 2008)).

⁴² See *Johnson & Johnson-Merck Consumer Pharm. Co. v. Rhone-Poulenc Rorer Pharm., Inc.*, 19 F.3d 125, 134 (3d Cir. 1994) (stating that "a well-designed consumer survey" asks "'comprehension' questions to determine what the viewers thought the message meant").

⁴³ See *Vision Sports*, 888 F.2d at 615 ("An expert survey of purchasers can provide the most persuasive evidence of secondary meaning."); 6 J. THOMAS MCCARTHY, MCCARTHY ON TRADEMARKS AND UNFAIR COMPETITION § 32:195 (4th ed. 2016) (explaining that plaintiff's failure to present a consumer survey creates the perception that it is "less than deadly serious about [] its case"); Joshua M. Dalton & Ilisa Horowitz, *Funny When You Think About It: Double Entendres and Trademark Protectibility*, 88 J. PAT. & TRADEMARK OFF. SOC'Y 649, 652 (2006) (explaining that surveys are "all but indispensable" in trademark litigation); Sandra Edelman, *Failure to Conduct a Survey in Trademark Infringement Cases: A Critique of the Adverse Inference*, 90 TRADEMARK REP. 746, 747 (2000) (explaining that "survey evidence has become de rigeur in trademark infringement cases" and "many courts will draw an adverse inference against a plaintiff . . . if a survey is not introduced").

⁴⁴ Shari Seidman Diamond & David J. Franklyn, *Trademark Surveys: An Undulating Path*, 92 TEX. L. REV. 2029, 2037 (2014).

A. Trademark and Unfair Competition Surveys

1. The Essential Role of Surveys

A major challenge for a court applying the survey interpretation method would be to determine the reliability of the survey results and the legal conclusions to be drawn from them. This is a challenge quite similar to the one courts have been wrestling with over several decades of trademark and unfair competition litigation. Thus, there is no need to reinvent the wheel. Let us examine how courts have applied and justified the method.

Survey evidence has several key applications in trademark litigation. One major factor in granting trademark protection is whether a mark, which consists of descriptive terms, has acquired secondary meaning.⁴⁵ And conversely, trademark protection may be lost if a mark that consists of fanciful terms has become generic.⁴⁶ Surveys are also dispositive when a court must decide whether a mark has been infringed.⁴⁷ In these inquiries, the question is whether a likelihood of consumer confusion exists.⁴⁸ Surveys address this question directly, asking people to report the inferences they make when seeing the mark.⁴⁹

Lawyers showing up in court without consumer survey evidence to back them may as well show up in a t-shirt and shorts, as “survey evidence has become de rigueur in trademark infringement cases.”⁵⁰ Although the Restatement on Unfair Competition counsels against this inference,⁵¹ “many courts will draw an adverse inference against a plaintiff . . . if a survey is not introduced.”⁵²

Survey evidence is regularly dispositive at the summary judgment stage.⁵³ Courts often grant summary judgment based on survey evi-

⁴⁵ MARY LAFRANCE, UNDERSTANDING TRADEMARK LAW 569 (2d ed. 2009).

⁴⁶ See Diamond & Franklyn, *supra* note 44, at 2032–40 (providing examples of marks becoming generic and losing protection).

⁴⁷ Robert H. Thornburg, *Trademark Survey Evidence: Review of Current Trends in the Ninth Circuit*, 21 SANTA CLARA COMPUTER & HIGH TECH. L.J. 715, 743 (2005) (stating that “well executed surveys, conducted in multiple locations, to the proper universe of consumers, is almost dispositive of many of the issues found in trademark disputes”).

⁴⁸ Diamond & Franklyn, *supra* note 44, at 2034–35.

⁴⁹ See Larry C. Jones, *Developing and Using Survey Evidence in Trademark Litigation*, 19 MEM. ST. U. L. REV. 471, 482 (1989) (providing that an acceptable survey once asked, as a follow-up, “What was there about the sign that made you say that?” (quoting Exxon Corp. v. Tex. Motor Exch. of Hous., Inc., 628 F.2d 500, 507 (5th Cir. 1980))).

⁵⁰ Edelman, *supra* note 43, at 747.

⁵¹ RESTATEMENT (THIRD) OF UNFAIR COMPETITION § 23 cmt. c (AM. LAW INST. 1995).

⁵² Edelman, *supra* note 43, at 747.

⁵³ Thornburg, *supra* note 47, at 743.

dence,⁵⁴ and they deny summary judgment motions when survey evidence shows an issue of material fact.⁵⁵ Defendants can succeed in offering exculpatory surveys,⁵⁶ and parties bearing the burden of proof can lose at the summary judgment stage if their survey evidence is flawed, even if the other side does not offer a survey.⁵⁷

Commentators have documented the central role of survey evidence in trademark litigation using various methodologies, particularly by looking at reported case outcomes.⁵⁸ But looking only at published opinions may create a biased impression of the impact of surveys.⁵⁹ An especially telling account comes from a study of trademark lawyers, who were asked to evaluate the impact of consumer surveys on cases that settled before summary judgment or a trial on the merits. It found “that surveys are used heavily in pretrial assessments and strategic decision making” and play “key roles in claim evaluation and are understood by attorneys as an influential settlement tool for both sides.”⁶⁰ Evidence of successful use at trial, therefore, “is just the tip of the iceberg” when it comes to the impact of surveys in shaping trademark outcomes.⁶¹

Survey evidence has similar crucial applications in other areas of unfair competition law, particularly in false advertising claims where the allegation is that consumers were misled or confused. For example, an allegation that the defendant’s advertisement implied falsity requires evidence about the mental impressions the ad left on con-

⁵⁴ See, e.g., *Big Island Candies, Inc. v. Cookie Corner*, 269 F. Supp. 2d 1236, 1250 (D. Haw. 2003) (acknowledging that judges often expect survey evidence when evaluating “genericness” issues).

⁵⁵ See, e.g., *Thane Int’l, Inc. v. Trek Bicycle Corp.*, 305 F.3d 894, 902 (9th Cir. 2002).

⁵⁶ *Fla. Int’l Univ. Bd. of Trs. v. Fla. Nat’l Univ., Inc.*, 830 F.3d 1242, 1266–67 (11th Cir. 2016) (“The district court reasonably credited FNU’s expert testimony that, while 30% to 50% of survey respondents thought that the State of Florida or some other government entity operates FNU, less than 1% of survey respondents associated FNU with FIU.”).

⁵⁷ See, e.g., *Spraying Sys. Co. v. Delavan, Inc.*, 975 F.2d 387, 394 (7th Cir. 1992) (affirming summary judgment by finding a consumer survey too flawed to create genuine issue of material fact).

⁵⁸ See, e.g., Dan Sarel & Howard Marmorstein, *The Effect of Consumer Surveys and Actual Confusion Evidence in Trademark Litigation: An Empirical Assessment*, 99 TRADEMARK REP. 1416, 1432 (2009); Itamar Simonson, *The Effect of Survey Method on Likelihood of Confusion Estimates: Conceptual Analysis and Empirical Test*, 83 TRADEMARK REP. 364, 366 (1993).

⁵⁹ Barton Beebe, *An Empirical Study of the Multifactor Tests for Trademark Infringement*, 94 CALIF. L. REV. 1581, 1641 (2006) (claiming that the impact of consumer surveys on litigation outcomes is overrated). For critiques, see, for example, Diamond & Franklyn, *supra* note 44, at 2043.

⁶⁰ Diamond & Franklyn, *supra* note 44, at 2061.

⁶¹ *Id.* at 2050.

sumers.⁶² A leading court established that “the success of a plaintiff’s implied falsity claim usually turns on the persuasiveness of a consumer survey.”⁶³

The importance of surveys is likely a product of both the effectiveness of survey evidence in supporting a litigant’s case and the signal that conducting a survey makes to the other party in demonstrating seriousness and belief in success.⁶⁴ This background suggests that the use of survey evidence is working to inform litigants about the viability of their cases. Cases settle in the shadow of expected trial outcomes; in trademark and unfair competition law, they settle in the shadow of the surveys.

2. *The Mechanics of Trademark and False Advertising Surveys*

Trademark law has a good deal to say about the evidentiary standards necessary for survey research to be admissible in court. Obviously, surveys have to be designed and conducted by experts, addressed to appropriate respondents, and ask non-leading questions.⁶⁵ Judges exclude surveys that fail to satisfy such basic reliability standards under Federal Rule of Evidence 403.⁶⁶ Even if admitted, the weight of survey evidence varies according to, among other factors, how well the survey captured the relevant “universe” of consumers,⁶⁷ how representative the sample from that universe was,⁶⁸ whether a

⁶² Johnson & Johnson * Merck Consumer Pharm. Co. v. Smithkline Beecham Corp., 960 F.2d 294, 297 (2d Cir. 1992).

⁶³ *Id.*; see also Novartis Consumer Health, Inc. v. Johnson & Johnson-Merck Consumer Pharm. Co., 290 F.3d 578, 588 (3d Cir. 2002) (“Novartis should have been required to prove through a consumer survey that the name and advertising actually misled or had a tendency to mislead consumers into believing that the product provided nighttime heartburn relief superior to any other product in the market.”).

⁶⁴ See Laura A. Heymann, *Surveying the Field: The Role of Surveys in Trademark Litigation*, JOTWELL (Feb. 23, 2015), <http://ip.jotwell.com/surveying-the-field-the-role-of-surveys-in-trademark-litigation/> (stating that a party is in a much better position to litigate or settle a trademark case when investing in consumer surveys).

⁶⁵ James Burrough Ltd. v. Sign of Beefeater, Inc., 540 F.2d 266, 278 (7th Cir. 1976); Union Carbide Corp. v. Ever-Ready Inc., 531 F.2d 366, 387–88 (7th Cir. 1976); Simon Prop. Grp. L.P. v. mySimon, Inc., 104 F. Supp. 2d 1033, 1038 (S.D. Ind. 2000).

⁶⁶ See Robert H. Thornburg, *Trademark Surveys: Development of Computer-Based Methods*, 4 J. MARSHALL REV. INTELL. PROP. L. 91, 93 (2004); see also J & J Snack Foods, Corp. v. Earthgrains Co., 220 F. Supp. 2d 358, 370 (D.N.J. 2002) (rejecting a survey because the attorneys confused descriptive with suggestive marks such that “the survey has no bearing on the issue it was submitted for”).

⁶⁷ MCCARTHY, *supra* note 43, at § 32:159 (“The first step in designing a survey is to determine the ‘universe’ to be studied . . . [.] [t]he segment of the population whose perceptions and state of mind are relevant to the issues in the case.”).

⁶⁸ See, e.g., Nabisco, Inc. v. PF Brands, Inc., 50 F. Supp. 2d 188, 211 (S.D.N.Y. 1999) (finding a survey flawed for failure to focus on the relevant “universe”).

control group was maintained to exclude confounding factors,⁶⁹ and whether the results of the survey were verified.⁷⁰ Courts examine whether the design, questionnaires, and interviews were unbiased, and they scrutinize the accuracy of the data analysis and report.⁷¹

The incentives to design reliable surveys depend on the reaction of courts to sloppy ones. Usually, minor errors result in the survey receiving diminished weight in any subsequent adjudication.⁷² But courts take a more aggressive approach towards survey results that are more prejudicial than probative, excluding such evidence.⁷³ Over time, courts have been converging on the appropriate trademark survey methodologies.⁷⁴ The Federal Judicial Center publishes a reference guide on survey research that was authored by one of the field's

⁶⁹ See, e.g., *Masterfoods USA v. Arcor USA, Inc.*, 230 F. Supp. 2d 302, 305 (W.D.N.Y. 2002) (using multiple controls to eliminate statistical noise).

⁷⁰ See, e.g., *Schieffelin & Co. v. Jack Co. of Boca*, 850 F. Supp. 232, 247 (S.D.N.Y. 1994) (showing the use of verification to cull respondents). For a summary of the Second Circuit's requirements for surveys, see *Toys "R" Us, Inc. v. Canarsie Kiddie Shop, Inc.*, 559 F. Supp. 1189, 1205 (E.D.N.Y. 1983).

⁷¹ See *J & J Snack Foods*, 220 F. Supp. 2d at 369 (summarizing Third Circuit requirements for survey reliability); *Exxon Corp. v. Xoil Energy Res., Inc.*, 552 F. Supp. 1008, 1022 n.20 (S.D.N.Y. 1981) ("The evidentiary value of plaintiff's survey is lessened considerably by the absence of practices and procedures . . . useful in assessing the validity of survey results[,] . . . including: use of only non-leading questions; verification by reinterviewing a substantial portion of those interviewed; design by qualified experts; and administration by impartial interviewers."). See generally McCARTHY, *supra* note 43, at § 32:171 (providing examples of methodological deficiencies).

⁷² See McCARTHY, *supra* note 43, at § 32:170 (stating that "technical deficiencies can reduce a survey's weight"); see also *Jellibeans, Inc. v. Skating Clubs of Ga., Inc.*, 716 F.2d 833, 844 (11th Cir. 1983) (citing *Exxon Corp. v. Tex. Motor Exch. of Hous., Inc.*, 628 F.2d 500, 507 (5th Cir. 1980)) (stating that "alleged technical deficiencies affect the survey's weight, however, and not its admissibility"); *THOIP v. Walt Disney Co.*, 690 F. Supp. 2d 218, 219 (S.D.N.Y. 2010) ("No survey is perfect and the limits and flaws of a survey generally go to evidentiary weight . . ."); *MasterCard Int'l Inc. v. First Nat'l Bank of Omaha, Inc.*, No. 02 Civ. 3691, 2004 WL 326708, at *7 (S.D.N.Y. Feb. 23, 2004) (stating that errors in the survey methodology go to the weight of the evidence); *Thornburg*, *supra* note 47, at 718 ("The Ninth Circuit's uniqueness with regard to trademark surveys is based largely upon its almost 'carte blanche' refusal to exclude survey evidence based upon technical deficiencies.").

⁷³ See, e.g., *GoSmile, Inc. v. Levine*, 769 F. Supp. 2d 630, 642–45 (S.D.N.Y. 2011) (questioning the probative value of surveys offered to establish consumer confusion). See generally Richard J. Leighton, *Using Daubert-Kumho Gatekeeping to Admit and Exclude Surveys in Lanham Act Advertising and Trademark Cases*, 92 TRADEMARK REP. 743, 747–53 (2002) (explaining the application of *Daubert-Kumho* to increasing challenges to the admission of survey evidence).

⁷⁴ See *Ass Armor, LLC v. Under Armour, Inc.*, No. 15-cv-20853, 2016 WL 7156092 at *2 (S.D. Fla. Dec. 8, 2016) (citing the factors advanced by the Judicial Center's Manual for Complex Litigation).

leading academics.⁷⁵ Likewise, the International Trademark Association (whose position should be taken with a grain of salt) concludes that “it is clear that the approach to the design, execution and presentation of an influential trademark survey is reasonably universal, perhaps due to its basis upon scientific principles” and that “common law jurisdictions . . . have substantially greater guidance available [for courts], reflecting a well-developed practice of using surveys as evidence.”⁷⁶ Few scholarly commentators doubt the efficacy of survey evidence per se, with the consensus being that⁷⁷ “the field of survey research incorporates all the essential structural techniques of other scientific expert evidence, including rigorous hypothesis testing, experimental design, control conditions, and statistical inference.”⁷⁸

There are occasional skeptical voices, and it might not surprise anyone that one of them came from Judge Posner, who referred to statistical surveys as a “black art[.]”⁷⁹ Writing in 1994, Judge Posner questioned whether there was a sound consensus in survey methodology, arguing that “[c]onsumer surveys conducted by party-hired expert witnesses are prone to bias” since “[t]here is *such* a wide choice of survey designs, none foolproof.”⁸⁰ A more polite version of this concern argues that courts’ treatment of surveys “has been plagued by inconsistencies and that courts need to more clearly elucidate ex ante rules” governing proper procedure.⁸¹ Some of the inconsistency and bias may be artifacts of the asymmetry of resources between the parties.

This is of course a valid concern, but it applies to all aspects and methods of litigation and does not uniquely or disproportionately con-

⁷⁵ See Shari Seidman Diamond, *Reference Guide on Survey Research*, in REFERENCE MANUAL ON SCIENTIFIC EVIDENCE 229 (2d ed. 2000), https://triacentral.clu-in.org/tech/documents/Fed_Jud_Center_Paper_on_Scientific_Evidence.pdf.

⁷⁶ COURTS & TRIBUNALS SUBCOMM., INT’L TRADEMARK ASS’N, REPORT ON BEST PRACTICES IN CONDUCTING SURVEYS IN TRADEMARK MATTERS 1 (2013), <http://www.inta.org/PDF%20Library/INTA%20Report%20on%20Best%20Practices%20in%20Conducting%20Surveys%20in%20Trademark%20Matters.pdf>.

⁷⁷ See Bird & Steckel, *supra* note 41, at 1026 (“Some commentators have similarly elevated consumer surveys to high evidentiary importance.”).

⁷⁸ Ruth M. Corbin & Arthur Renaud, *What’s Good for the Goose Is Bad for the Gander: Why Confusion Surveys for Plaintiff and Defendant Should Be Different*, 16 INTELL. PROP. J. 177, 180 (2003).

⁷⁹ *Indianapolis Colts, Inc. v. Metro. Balt. Football Club Ltd. P’ship*, 34 F.3d 410, 416 (7th Cir. 1994).

⁸⁰ *Kraft Foods Grp. Brands LLC v. Cracker Barrel Old Country Store, Inc.*, 735 F.3d 735, 741 (7th Cir. 2013).

⁸¹ Irina D. Manta, *In Search of Validity: A New Model for the Content and Procedural Treatment of Trademark Infringement Surveys*, 24 CARDOZO ARTS & ENT. L.J. 1027, 1032, 1062–66 (2007) (demonstrating some lack of similarity between the survey conditions and marketplace conditions).

demn the survey methodology. In light of such concerns, a large background of law is already in place making survey wording and technique in trademark disputes reasonably consistent across cases.⁸² The conclusion is that “[a] substantial amount of case law exists which provides insight into how to conduct and prepare a trademark survey that will be admissible in court.”⁸³ Indeed, fifteen years after his “black art” characterization, Judge Posner had changed his tune, writing (in a false advertising case) that to determine whether a statement is misleading, “the best evidence is a responsible survey.”⁸⁴

Trademark surveys are certainly imperfect, but the consensus is that they are preferable on balance to alternative methods for establishing the effects of marketing messages and the likelihood of consumer confusion. A quick examination of the case law suggests that the survey skeptics’ arguments have not carried the day.⁸⁵ Surveys that speak to likelihood of confusion use several standardized formats,⁸⁶ approved by federal appellate courts.⁸⁷ Similarly, surveys establishing whether a mark is generic are subject to court-imposed methodological guidelines.⁸⁸ Courts are paying attention to the design of surveys and the suitability of various methodologies, just as they do with other kinds of expert evidence.

To be sure, slight differences in survey parameters such as question wording and sampling can have decisive effects on survey evidence.⁸⁹ Surveys can be manipulated so as to elicit the desired results; they also can be resampled or selectively presented.⁹⁰ Even academic surveys are prone to this defect, presenting difficulty for the research

⁸² See Jerre B. Swann, *Judge Richard Posner and Consumer Surveys*, 104 TRADEMARK REP. 918, 921 (2014) (providing examples of courts keeping manipulation of consumer surveys by experts under control).

⁸³ Thornburg, *supra* note 66, at 91.

⁸⁴ *Muha v. Encore Receivable Mgmt., Inc.*, 558 F.3d 623, 628 (7th Cir. 2009).

⁸⁵ See *Simon Prop. Grp. L.P. v. mySimon, Inc.*, 104 F. Supp. 2d 1033, 1038 (S.D. Ind. 2000) (“Several models for consumer surveys on likelihood of confusion have found acceptance in the courts.”).

⁸⁶ See INT’L TRADEMARK ASS’N, *supra* note 76, at 13.

⁸⁷ For a framework set by the Seventh Circuit for the presentation of survey questions, see *Union Carbide Corp. v. Ever-Ready, Inc.*, 531 F.2d 366, 386–87 (7th Cir. 1976). See also *Simon Prop. Grp. L.P.*, 104 F. Supp. 2d at 1038 (explaining the suitability of the *Union Carbide* framework for the question asked and rejecting the survey provided for failing to answer the question presented as effectively). For a framework set by the Eighth Circuit, see *SquirtCo. v. Seven-Up Co.*, 628 F.2d 1086, 1091 (8th Cir. 1980).

⁸⁸ See *J & J Snack Foods, Corp. v. Earthgrains Co.*, 220 F. Supp. 2d 358, 369–72 (D.N.J. 2002) (criticizing the survey in question for confusing the correct definition of law, failing to correctly identify the relevant universe, and failing to supply sufficient information to survey takers).

⁸⁹ See, e.g., *Simonson*, *supra* note 58, at 365–66.

⁹⁰ *Kraft Foods Grp. Brands LLC v. Cracker Barrel Old Country Store, Inc.*, 735 F.3d 735, 741 (7th Cir. 2013) (stating that “[c]onsumer surveys conducted by party-hired expert

community in assessing the data practices underlying surveys' findings.

This problem is significant, but it is not unique to survey evidence. Much of the evidence presented to courts in adversarial proceedings by parties interested in affecting the result is cherry-picked. There is a large literature on these choices, supported by science on consumer behavior and psychology, to help guide courts and litigants as to the appropriate approach.⁹¹ The solutions typically applied to such credibility challenges could provide some assurance, and we discuss them in Section II.B below.

It is telling that the judicial enthusiasm for the survey method in trademark litigation has not subsided, despite the real challenges. The inadequacy of alternative tools of interpretation explains the increasing prevalence of the survey method. Surveys have displaced evidence that is manifestly worse at evaluating whether consumers were likely to be confused. Without survey evidence, courts are asked to either evaluate the testimony of individual members of the public, who are exceedingly unlikely to show “a fairly representative picture,”⁹² or to engage in “an exercise in pure judicial fantasy”⁹³ in speculating how consumers likely understand the communication. Judges are asked to evaluate how people with lesser education and legal experience appreciate marketing messages, and they are sometimes explicit in conceding that without survey evidence they are unlikely to reach proper conclusions.⁹⁴ Trademark and unfair competition law ask a variety of empirical questions and are increasingly using empirical tools to answer them.

witnesses are prone to bias” since “[t]here is *such* a wide choice of survey designs, none foolproof”).

⁹¹ See Swann, *supra* note 82, at 920–22 (describing examples of cognitive psychology research that can guide courts in analyzing consumer surveys).

⁹² Premier-Pabst Corp. v. Elm City Brewing Co., 9 F. Supp. 754, 760 (D. Conn. 1935).

⁹³ Sun Life Assurance Co. of Can. v. Sunlife Juice Ltd. (1988), 65 O.R.2d 496, 500 (Can. Ont. H. Ct. J.).

⁹⁴ Judge Weinstein was particularly candid on this issue, stating that:

“[A] federal trial judge, with a background and experience unlike that of most consumers, is hardly in a position to declare, ‘Because I appreciate that the television campaign is just expressing a far-fetched opinion and not making a statement of fact, all viewers must appreciate it as well.’ . . . Arguably, the communication . . . [may] mean something quite different to the viewer. This central issue cannot be resolved without surveys, expert testimony, and other evidence of what is happening in the real world of television watchers.”

Verizon Directories Corp. v. Yellow Book USA, Inc., 309 F. Supp. 2d 401, 407 (E.D.N.Y. 2004).

B. *The Survey Method in Contract Interpretation*

I. *Whom to Survey?*

The benchmark case for the application of the survey interpretation method is a consumer contract. Like advertisements and trademarks, the provisions of consumer contracts are directed at consumers, and in interpreting these provisions courts are trying to determine how consumers likely understand them. Current interpretation doctrine asks courts to speculate about the answer to this question. The survey interpretation method directs courts to rely instead on the opinion of large samples of respondents who resemble the contract's audience.

Interpreting consumer contracts through surveys requires a sample of like consumers. It is fairly straightforward to identify the demographics and geographical concentrations of the consumers governed by the contract and construct a representative sample. Obviously, respondents need not come solely from consumers who actually accepted this contract.

The survey interpretation method is potentially well suited to other contracting environments, like merchant-to-merchant contracts and negotiated agreements between firms. In these settings, the universe of potential survey respondents has to be adjusted to include merchants, lawyers, and other professionals—respondents who have more insight as to the meaning of the contract. It might be harder to recruit neutral respondents in some of these contexts because people familiar with the sector may have a stake in how the interpretive battle is resolved, which means that the size of the survey would have to be smaller. Nevertheless, when sector knowledge is necessary, there is all the more reason to outsource the interpretation to knowledgeable insiders.⁹⁵

Consider, for example, the interpretation of sovereign bond contracts. Like consumer contracts, they are all boilerplate. But unlike consumer contracts, they govern transactions with investors, many of whom are sophisticated. Their terms are notoriously subject to lengthy and costly interpretation disputes.⁹⁶ Could they be interpreted, instead, by surveys of practitioners? It turns out that the existence of

⁹⁵ John F. Coyle, *The Canons of Construction for Choice-of-Law Clauses*, 92 WASH. L. REV. 631, 682–87 (2017) (showing that in the absence of a systematic survey, judges can interpret contract language in ways that conflict with the parties' intentions); *see also* Solan et al., *supra* note 39, at 1292 (reporting on systematic differences in the way that judges and lay respondents interpreted language in contract disputes).

⁹⁶ *See generally* MITU GULATI & ROBERT E. SCOTT, *THE THREE AND A HALF MINUTE TRANSACTION* 19–28 (2013); Stephen Choi et al., *The Black Hole Problem in Commercial Boilerplate* (N.Y.U. Sch. of Law, Public Law Research Paper No. 16–40, 2016).

interpretation risk has led the Bank of England and top law firms to form committees in which representative groups of practitioners are asked to interpret sovereign bond contract language, with an eye to helping courts reach more objective interpretations.⁹⁷ While the reports are far more detailed and technical than the surveys we imagine (in part because they do not focus on specific disputes), they nevertheless prove the existence of groups of potential survey respondents with sufficient expertise to whom the interpretation may be outsourced. Similarly, in interpreting negotiated contracts, the best group to sample may be transactional lawyers.⁹⁸

While consumer contracts provide the most straightforward application of the survey interpretation method, there are significant advantages to interpretive consistency across contract types. Using surveys to interpret all contracts takes some pressure off the determinations that courts otherwise would have to make in boundary cases. Some discretion would be necessary, at least early on, as to which contracts are most susceptible to survey interpretations, and the uncertainty about how those judgments would be resolved will impose costs on the parties. The relevant question, of course, is how these costs compare to those engendered by the substantial uncertainty that presently exists in contract interpretation.

2. *What to Ask?*

If the survey produces a winner—an interpretation supported by a statistically significant and large enough majority—the court would adopt this as the meaning of the disputed language. If no clear winner emerges, the court would rule that the disputed term is inherently ambiguous and the party that bears the burden to prove a specific meaning of the term loses. Either way, the survey method absolves the court from the agonizing law/fact dichotomy that besets interpretation doctrine. Instead of a judge making the pretrial (legal) decision whether the language has unambiguous meaning (and, if ambiguous, a trial of fact ensues), the survey resolves both issues of law and fact conclusively. Survey respondents are both the judge and (a very large) jury. Indeed, under the survey interpretation method the winner ought to prevail on a motion for summary judgment.

⁹⁷ See, e.g., FIN. MKTS. LAW COMM., *Pari Passu* Clauses: Analysis of the Role, Use and Meaning of *Pari Passu* Clauses in Sovereign Debt Obligations as a Matter of English Law (2015), http://www.fmlc.org/uploads/2/6/5/8/26584807/fmlc_paper_analysing_the_role_use_and_meaning_of_pari_passu_clauses_in_sovereign_debt_obligations_as_a_matter_of_english_law.pdf.

⁹⁸ We consider some of the details of how that might work *infra* Section IV.E.

In theory, survey respondents could be shown any number of alleged facts that might influence the meaning they assign to the disputed language. The method is thus agnostic with respect to the major debate within contract interpretation doctrine—the Williston/Corbin divide on text versus context.⁹⁹ As contract formalists would like it, respondents could be shown only the disputed text; or, as the realists advocate, they could be exposed to additional facts surrounding the case. Practically, however, the survey method—by typically relying on respondents with limited attention and sophistication—restricts the quantum of such background facts. It thus relies on an interpretation that derives from exposure to, at most, very limited context.

The question of how much information to give the survey respondents is an important methodological challenge in the design of contract interpretation surveys, which goes beyond the issues raised by trademark surveys. Asking respondents to read the entire contract will almost never be the right answer because survey respondents have limited attention spans just like ordinary consumers, and the more they are asked to read, the less sure we can be that they have done so diligently. The longer the text shown to respondents, the more likely it becomes that respondents will answer based on their normative preferences or provide answers at random.

This is not the only challenge. Another question is whether respondents should be shown only the plain text and the two proposed interpretations, or whether they should review the best arguments of each side in interpreting the language. If they see only the plain text, how much of it (and of the surrounding clauses that might affect its intended meaning) should respondents see? A stripped-down survey more closely resembles how lay readers approach contracts, but the alternative approach is defensible and likely desirable for some business contracts. As in trademark litigation, the method would likely evolve over time to reflect both methodological rigor and pragmatic constraints. It is perhaps more important that such design questions be resolved definitively than correctly.¹⁰⁰

⁹⁹ See, e.g., FARNSWORTH, *supra* note 11, at 422 (summarizing the tension between the four corners test and the concept that all circumstances should be considered when determining the meaning of a contract); ROBERT A. HILLMAN, *THE PRINCIPLES OF CONTRACT LAW* 232–37 (2004) (explaining the role that the parol evidence rule plays in contract interpretation).

¹⁰⁰ Cf. David A. Strauss, *Common Law Constitutional Interpretation*, 63 U. CHI. L. REV. 877, 879–80 (1996) (“[S]ometimes it is more important for a matter to be settled than for it to be settled right.”).

3. *How to Determine a Winner?*

A host of technical questions also arises in implementing survey techniques. What ratio of prevalence is necessary? For example, does a 51%–49% margin suffice, assuming the sample size is large enough to render that result a statistically significant difference from a 50%–50% split? Should courts count differently respondents who say that the language “definitely” means *X* versus those who lean in that direction but say that they are less sure? We think there is a region of outcomes around 50%–50% that are close enough to a tie that the meaning of the term ought to be viewed as inherently ambiguous. In such cases, again, the party that bears the burden to prove a specific meaning of the term loses. In Part III below, we utilize insights from the interpretation surveys we conducted to further discuss the proper majority thresholds.

Further, how should neutral votes count? If a large fraction of the respondents find the language at issue ambiguous, would it suffice for an interpretation to receive a majority among the remaining decisive respondents? In trademark law, some courts apply a “15% rule”—holding that consumer confusion exists if more than 15% of surveyed consumers are confused by the mark.¹⁰¹ This is a very low threshold, suitable perhaps to the protection of a proprietary mark, but not to the interpretation of contractual language. As we will discuss below,¹⁰² a sizeable minority of even well-compensated survey respondents will answer at random when facing questions with answers they have not previously considered. Given this propensity, we would be very reluctant to characterize contractual language as ambiguous just because 15% or 20% of a representative sample regard it as such.

In ordinary circumstances a claimant would have to demonstrate some clear majority to prevail. But the survey interpretation method is well suited to accommodate various strengths of legal presumptions that would determine the requisite majority. In some contexts, the majority threshold ought to be shifted against particular parties. For example, under the doctrine of *contra proferentem* interpretation (applicable primarily in insurance contract law), courts are instructed

¹⁰¹ *Exxon Corp. v. Tex. Motor Exch. of Hous., Inc.*, 628 F.2d 500, 507 (5th Cir. 1980) (noting that 15% of people surveyed confused the mark “Texon” with “Exxon”); Daniel Gervais & Julie M. Latsko, *Who Cares About the 85 Percent? Reconsidering Survey Evidence of Online Confusion in Trademark Cases*, 96 J. PAT. & TRADEMARK OFF. SOC’Y 265, 267 (2014). Thresholds on the order of 20% are prevalent in false advertising cases. See, e.g., *Procter & Gamble Co. v. Ultreo, Inc.*, 574 F. Supp. 2d 339, 345 (S.D.N.Y. 2008).

¹⁰² See *infra* Section III.B.3.

to interpret ambiguous terms against the drafter.¹⁰³ This principle can be implemented within the survey method by requiring the drafting party to achieve a “supermajority” of respondents’ support. A 60-40 or 67-33 split, for example, would be adequate for a court to accept the contract drafters’ preferred interpretation, but a 51-49 split would not be.¹⁰⁴ In general, courts could adjust the majority threshold to achieve any number of policy goals. If, for example, the law seeks to promote use of lay or compact language in contracts, a term that fails these standards could be “sanctioned” by having to achieve a supermajority.

An implicit question in applying the presumption against the drafter (or other interpretive presumptions) is how strong the presumption should be. In the context of the survey method, the question is what majority support the drafter would be required to show. In answering this question, the survey method provides an additional tool currently unavailable. As we show in Part III, the survey methodology can identify *alternative* formulations of language that achieve the drafter’s intended meaning with less ambiguity. If such formulations were available but not employed, the presumption against the drafter ought to be strengthened.

4. *Reliability of Survey Evidence*

An additional methodological challenge for courts would be to police biases in the execution of the survey experiment that can result from asymmetric resources of the parties or from manipulations by experts conducting the surveys. We think that courts can conquer this challenge. First, courts could rely on each party to scrutinize the

¹⁰³ See *Gaunt v. John Hancock Mut. Life Ins. Co.*, 160 F.2d 599, 602 (2d Cir. 1947) (“[C]ontra proferentem is more rigorously applied in insurance than in other contracts, in recognition of the difference between the parties in their acquaintance with the subject matter.”); 5 MARGARET N. KNIFFIN, *CORBIN ON CONTRACTS: INTERPRETATION OF CONTRACTS* § 24.27 (Joseph M. Perillo ed., rev. ed. 1998) (explaining that contra proferentem is a “technique” in which courts “adopt the meaning that is less favorable in its legal effect to the party who chose the words”).

¹⁰⁴ Here we part ways with Boardman, *supra* note 39, at 1111–12, who treats a situation where 65% of respondents accept one interpretation and 25% accept a different interpretation as an instance of ambiguity. Boardman’s paper is theoretical—she did not run experiments like the ones we conducted. Our data suggests that if the threshold for ambiguity is set as high as she suggests, then nearly every contract will be deemed ambiguous. Our data shows that survey respondents in general are highly heterogeneous, and it is difficult to prevent the least conscientious among them from answering more or less at random or ignoring the text of the contract and interpreting entirely on the basis of the context and their priors. Those facts do not discredit the methodology; they merely require careful consideration of the appropriate thresholds. See *generally id.* at 1116 (“The consumer-research joke that one can find 10% of people who believe anything is funny because it’s true.”).

survey evidence produced by the other side and to highlight any defects.¹⁰⁵ The survey method is readily amenable to experimental interventions. Subjects can be randomly assigned to different treatments (e.g., prompts, sequences of questions, phraseologies), thereby enabling survey researchers to isolate the effects of particular survey design choices on substantive responses. Very few other forms of evidence introduced in litigation have that reassuring attribute.¹⁰⁶

Second, courts could threaten to completely disregard surveys that are tainted in order to induce parties to provide competent ones.¹⁰⁷ Pushing the threat further, courts could deploy a method similar to final offer arbitration: In choosing between the two competing surveys presented by the litigants, courts could rely entirely on the one that they perceive to have followed more reasonable protocols. This procedure has well-documented effects of moderating the parties' self-serving positions.¹⁰⁸ The benefit to each party of designing a biased survey is at least partially offset by the risk of having the survey "defeated" by the other party's more reasonable design.¹⁰⁹

Third, courts could rely on court-appointed experts to evaluate the credibility of surveys presented by the parties. Indeed, federal courts already do precisely that.¹¹⁰ If parties cannot be trusted to elicit reliable surveys, the method could be restricted to surveys conducted by neutral experts, initiated and solicited by courts on behalf of both parties.¹¹¹ The procedural authority to do so, of course, exists.¹¹² In

¹⁰⁵ See Corbin & Renaud, *supra* note 78, at 178 (explaining the strategies a party can use to demonstrate weaknesses in the opposing party's surveys).

¹⁰⁶ For a discussion of how this advantage provides insights not otherwise available, see *infra* Part III.

¹⁰⁷ See, e.g., *Malletier v. Dooney & Bourke, Inc.*, 525 F. Supp. 2d 558, 563 (S.D.N.Y. 2007) (deciding to follow special master academics' recommendations to exclude several expert surveys on the basis of methodological flaws).

¹⁰⁸ See B. Jay Coleman et al., *Convergence or Divergence in Final-Offer Arbitration in Professional Baseball*, 31 INDUS. REL. 238, 244 (1993) (analyzing the effect of final-offer arbitration on the positions parties present to the tribunal).

¹⁰⁹ This type of Final Offer Mechanism can be applied to other problems of contract interpretation. See Omri Ben-Shahar et al., *An Ex-Ante View of the Battle of the Forms: Inducing Parties to Draft Reasonable Terms*, 25 INT'L REV. L. & ECON. 350, 353 (2005) (discussing ways for courts to select the most reasonable standard terms in battle of the forms litigation, including the final-offer arbitration mechanism).

¹¹⁰ See, e.g., *Calvin Klein Cosmetics Corp. v. Parfums de Coeur Ltd.*, 824 F.2d 665, 670–72 (8th Cir. 1987) (affirming trial court's appointment of a special master to evaluate survey and other evidence in a trademark dispute); *Malletier*, 525 F. Supp. 2d at 563 (following special master academics' recommendations to exclude several expert surveys on the basis of methodological flaws); *SunAmerica Corp. v. Sun Life Assurance Co. of Can.*, 890 F. Supp. 1559, 1570–71 (N.D. Ga. 1994) (involving joint survey conducted by the litigants at the court's suggestion), *vacated on other grounds*, 77 F.3d 1325 (11th Cir. 1996).

¹¹¹ Phyllis J. Welter, *A Call to Improve Trademark Survey Evidence*, 85 TRADEMARK REP. 205, 209 (1995) (calling for the administration of surveys by court-appointed neutral

anticipation, the parties may agree that the court would select the survey firm. Or, the parties may contract into an alternative dispute resolution process that is affiliated with a neutral survey firm. Arbitrators, for example, may offer to the parties a forum that employs, alongside the legally trained arbitrator, a survey methodologist who enables the arbitrator to put contested contractual language to the test of empirical surveys. In this environment, one can imagine the emergence of firms specializing in consumer survey research with a reputation to protect. We discuss later, in Part V, how survey firms might branch out and also help parties ascertain the meaning of contracts at the time of drafting. Such profitable opportunities may bolster, rather than blur, the incentive of survey professionals to produce reliable evidence.

III

THE SURVEY INTERPRETATION METHOD IN ACTION

Part I of this Article explained the problem with existing contract interpretation doctrine. Part II offered to solve the problem by proposing that courts rely on large representative surveys to interpret contested language in contracts. In the remainder of this Article, we offer three types of support for the proposed survey interpretation method. Part III demonstrates the pragmatic argument: Surveys are practical and inexpensive. Later, Part IV will develop the fuller normative case, and Part V the doctrinal support, for the survey interpretation method.

To demonstrate the practical value of the survey interpretation method, we selected several cases in which courts had to interpret disputed language in contracts. We wanted to study how survey respondents interpret these contracts, how they perform vis-à-vis courts, and whether we can detect patterns that would give us confidence about entrusting the interpretive task to such crowds. We presented the essential facts of the chosen cases to large representative groups of respondents and asked them to interpret the contractual language. In this Part, we explain the methodology we used and the results we obtained. We contrast the results with the outcomes reached by courts.

experts); *see also* *Indianapolis Colts, Inc. v. Metro. Balt. Football Club Ltd. P'ship*, 34 F.3d 410, 415 (7th Cir. 1994) (noting that the survey method "might be improved by asking each party's hired expert to designate a third, a neutral expert who would be appointed by the court to conduct the necessary studies").

¹¹² FED. R. EVID. 706 (allowing courts to appoint expert witnesses).

A. Methodology

In designing the interpretation surveys, we needed to do more than throw facts of cases in front of respondents and ask them to vote. This would have proven nothing, since opinion polls can be done for every conceivable topic. Instead, what we hope to learn is whether the results of such surveys are consistent with essential patterns of sound contract interpretation. While our goal was not to develop the best practices for the design of interpretation surveys, we did hope to examine at least some baseline strategies for eliciting reliable responses.

We conducted the surveys in two waves, each containing several prompts based on the facts of litigated cases. Each prompt described, in one paragraph, the essence of the contested issue and produced the text of the disputed term in the contract. The order in which respondents saw these vignettes was randomized. The names of the parties were changed from those in the real cases. In no instance were the respondents given any information about how the courts had decided the disputes in question.

Respondents were asked to select between the interpretation of the contract offered by a plaintiff and a defendant. Specifically, they were given five possible answers in each case:¹¹³

- Plaintiff's argument about the contract's meaning is definitely right;
- Plaintiff's argument about the contract's meaning is probably right;
- It is completely uncertain whether the plaintiff's or defendant's argument about the contract's meaning is right;
- Defendant's argument about the contract's meaning is probably right;
- Defendant's argument about the contract's meaning is definitely right.

Wave 1 of the survey presented each respondent with the same vignettes. Wave 2 of the survey randomly assigned half the respondents to read slightly modified versions of most vignettes. In Wave 2, we were trying to determine how subjects would react if legally relevant or legally irrelevant changes were made to the vignettes.

Wave 2 asked respondents to answer not only how they interpreted the contract at issue, but also how the average person would interpret it. We did that to see whether respondents as a group were

¹¹³ We randomized the order of whether the continuum began with the plaintiff being definitely right, or with the defendant being definitely right. The randomly assigned order was held consistent throughout the survey for each respondent.

good at predicting how the majority of people would respond. If people can predict accurately how most of their peers would resolve a case involving contractual ambiguity, it strengthens the case for making survey evidence of this kind legally dispositive.

The experiment was administered online to a nationally representative sample recruited by Toluna, a well-regarded survey research firm.¹¹⁴ One advantage of the Toluna sample is that it is used relatively rarely by academic researchers, rendering the sample less polluted for our purposes.¹¹⁵ Wave 1 of the sample surveyed 1300 respondents, and Wave 2 surveyed 1294 respondents. In each wave, a few responses were discarded based on respondents completing the survey unusually quickly, and other attention checks were employed to make it likely that the respondents in the sample were reading the questions carefully. Thus, despite some minor differences between the two samples, both nicely reflected the demographics of the U.S. population as a whole, or at least the overwhelming majority of the population that uses the Internet.¹¹⁶

B. Results

1. Experiment #1: Ambiguous Homeowner's Insurance

The first experiment set out to test how respondents react to what many courts regarded as truly ambiguous language. A necessary condition for the reliability of the survey interpretation method is the *neutrality* of the respondents—that is, that ambiguous language produces a tie.

¹¹⁴ *About Us*, TOLUNA, <http://www.toluna-group.com/about-toluna/about> (last visited Apr. 21, 2017).

¹¹⁵ In Wave 1, the mean age of the 1300 respondents was 45.7 with a range of 18 to 90 years old and a standard deviation of 16.1. Females comprised 51.0% of the sample. In this sample, 81.5% of respondents self-identified as White, 10.5% as Black, and 3.5% as South or East Asian. On a separate question, 15.5% of the sample reported that they are Latino or Hispanic. About 11.4% of the sample had not finished high school, 30.4% had high school diplomas, 29.1% had some college experience, 19% had college degrees, and 10.2% had some kind of graduate degree.

Wave 2 of the sample was demographically similar. Among the 1294 respondents who satisfied our attention checks, the mean age was 46.8, with a range of 18 to 87 years old and a standard deviation of 16.94. Females comprised 52.6% of the sample. In this sample, 79.2% of respondents self-identified as White, 11.8% as Black, and 2.9% as South or East Asian. On a separate question, 17.2% of the sample reported that they are Latino or Hispanic. About 11.0% of the sample had not finished high school, 32.6% had high school diplomas, 29.1% had some college experience, 18.1% had college degrees, and 9.2% had some kind of graduate degree.

¹¹⁶ For a discussion of the demographics of internet and non-internet users in the United States, see Andrew Perrin & Maeve Duggan, *Americans' Internet Access: 2000–2015*, PEW RESEARCH CENTER (June 26, 2015), <http://www.pewinternet.org/2015/06/26/americans-internet-access-2000-2015/>.

We chose a term in insurance contracts that has long split courts, and which has often been characterized by courts as ambiguous.¹¹⁷ The courts that view the term as unambiguous are almost evenly split as to its meaning.¹¹⁸ The term deals, not surprisingly, with insurance coverage exclusions (often phrased as double- or triple-negative). Specifically, standard homeowners insurance policies contain coverage against tort liability arising from injuries to house visitors.¹¹⁹ These policies explicitly exclude injuries arising from commercial or business pursuits (coverage for these is sold separately, for higher premiums). Under the business-pursuits exclusion, for example, a doctor who runs his clinic from home is not covered by his homeowners insurance for malpractice liability.

But what about a babysitter? This question came up, for example, in *State Farm Fire & Casualty Co. v. Moore*,¹²⁰ in which the homeowner, Rebecca Moore, was watching her own child and was paid by her neighbors to also care for their son. The neighbor's son was injured when boiling water from the kitchen stove accidentally spilled on him. The neighbors sued, and Moore asked State Farm to defend her under the homeowners policy. State Farm refused, based on the policy's business-pursuits exclusion, which says: "This policy does not apply . . . to bodily injury or property damage arising out of business pursuits of any insured except activities therein which are ordinarily incident to nonbusiness pursuits."¹²¹

The trial court in Illinois granted summary judgment in State Farm's favor, interpreting the language to unambiguously exclude the injuries.¹²² The appellate court reversed. Focusing on the phrase "except activities . . . ordinarily incident to nonbusiness pursuits," two judges held that this exception to the exclusion applied unambiguously because boiling a pot of water was something Moore would have done for nonbusiness reasons.¹²³ They also cited the many prior cases that interpreted this exact phrase and found that courts reached conflicting holdings.¹²⁴ This suggested to the majority judges that, at the

¹¹⁷ See *Stanley v. Am. Fire & Cas. Co.*, 361 So. 2d 1030, 1032 (Ala. 1978); *Crane v. State Farm Fire & Cas. Co.*, 485 P.2d 1129, 1131 (Cal. 1971); *Nationwide Mut. Fire Ins. Co. v. Collins*, 222 S.E.2d 828, 830–31 (Ga. Ct. App. 1975).

¹¹⁸ See *Robinson v. Utica Mut. Ins. Co.*, 585 S.W.2d 593, 595–98 (Tenn. 1979) (citing numerous cases that construe the same policy language, with some finding it ambiguous and others not).

¹¹⁹ See, e.g., *Am. Family Mut. Ins. Co. v. Williams*, 832 F.3d 645, 647 (7th Cir. 2016).

¹²⁰ 430 N.E.2d 641 (Ill. App. 1981).

¹²¹ *Id.* at 643.

¹²² *Id.*

¹²³ *Id.* at 643, 646–47.

¹²⁴ *Id.* at 646–47.

very least, the language was ambiguous in its application to situations like Moore's.¹²⁵ And when insurance contract language is ambiguous, courts interpret it against the insurer.¹²⁶ The dissenting judge sided with the trial court, thinking that the exclusion applied unambiguously.¹²⁷

When we presented the respondents with the facts of *Moore* in Wave 1 of our study, we received the following results:¹²⁸

TABLE 1A. HOMEOWNERS INSURANCE

<i>Answers</i>	<i>Frequency</i>	<i>Percent</i>
Injury DEFINITELY covered	159	15.9
Injury PROBABLY covered	201	20.1
Coverage is COMPLETELY UNCERTAIN	182	18.2
Injury PROBABLY NOT covered	269	26.9
Injury DEFINITELY NOT covered	189	18.9
Total	1000	100.0

Respondents preferred the insurance company's interpretation, although not the overwhelming majority. 36% of the respondents said that the policy either definitely or probably covered the injuries, whereas almost 46% said that the policy definitely or probably did not cover the injuries. We confirmed that this pro-insurer interpretation is statistically significant.¹²⁹ Notice that in either the "covered" or the "uncovered" group a similar fraction of respondents was in the "definitely" category. Thus, in this particular case our results would not change if we weight more heavily the "definitely" responses.

While the insurer's interpretation prevailed, the margin was slim: Only 56% of those choosing one of the two sides favored the insurer. There was also a sizeable group of "uncertain," such that 54% of the respondents did not affirmatively choose the insurer's position. This

¹²⁵ *Id.* at 647 ("Under the terms of the clause, the particular facts of this case could reasonably be interpreted to be covered or excluded. Because of this ambiguity, doubt is resolved in favor of coverage for the Moores.").

¹²⁶ RESTATEMENT OF THE LAW OF LIABILITY INSURANCE § 4 (Proposed Final Draft 2017).

¹²⁷ 430 N.E.2d at 648 (Reinhard, J., dissenting).

¹²⁸ We removed from the results "implausible" answers—those who spent less than five minutes answering the questions.

¹²⁹ We confirmed this conclusion with the following statistical approach. Looking only at respondents who did not choose "uncertain," 44% of the remaining chose "covered" and 56% chose "not covered." Given that the data we observed is a random draw out of the population, how certainly can we accept the hypothesis that, in the actual population, less than 50% would respond "covered" (i.e., the insurance company prevails)? The likelihood that a majority of the population would favor "not covered" is, according to the test we ran, 99.97%—well above the conventional 95% acceptable confidence rate.

slim victory can be regarded as a weak confirmation for the results courts reach. We noted that courts are split on the question whether a babysitter is covered under the policy, either concluding that the policy is ambiguous or splitting fairly evenly as to its unambiguous meaning. Our respondents were similarly split: We cannot say that the insurer won a resounding survey victory. With such small margin, courts' instinct to interpret against the drafter could make good sense.

Given this less than ringing victory for the insurer's interpretation, we now wanted to test another important property: How do survey respondents change their interpretation when the ambiguity is reduced? For the method to be reliable, respondents must show a propensity to shift towards the position that the new language favors. To that end, we split the Wave 2 respondents into two groups. Half of them (the control group) were asked exactly the same question as the Wave 1 respondents. The other half (the treatment group) were shown shorter, seemingly less ambiguous and more pro-insurer, contractual language:

"This policy does not apply to bodily injury arising out of business pursuits of the homeowner."

We expected this language to have two effects: First, without the complex and confusing exception to the exclusion about "activities ordinarily incident to nonbusiness pursuits," fewer respondents would choose "uncertain." Second, without the exception to the exclusion, the exclusion could only be broadened, and therefore more respondents would side with the insurance company. We received the following results:

TABLE 1B. REDUCED AMBIGUITY HOMEOWNERS INSURANCE

<i>Answers</i>	Original Exclusion (Control Group)		Shortened Exclusion (Treatment Group)	
	<i>Frequency</i>	<i>Percent</i>	<i>Frequency</i>	<i>Percent</i>
Injury DEFINITELY covered	98	15.2	82	12.6
Injury PROBABLY covered	121	18.8	98	15.1
UNCERTAIN	122	18.9	96	14.8
Injury PROBABLY NOT covered	153	23.8	187	28.8
Injury DEFINITELY NOT covered	150	23.3	186	28.7
Total	644	100.0	649	100.0

Sure enough, respondents in the treatment group who were shown the new relatively unambiguous version of the policy overwhelmingly sided with the insurer's interpretation of no-coverage: 58% of the sample sided with the insurer (compared to 47% in the

control group), and only 28% sided with the policyholder (compared to 34% in the control group).¹³⁰ The percentage of the sample saying the policy's meaning was uncertain also declined (from 18% to 15%). This tells us that respondents as a whole reliably pick up on improvements in clarity. It also tells us that if the survey interpretation method were employed, it would have been possible for an insurance company like State Farm to draft its policy in a way that would cause a supermajority of respondents (more than 2-to-1) to embrace the company's preferred interpretation.

In sum, Experiment #1 provided some baseline assurance that ambiguous terms are viewed as such by survey respondents, and that improvements in the clarity of the terms correctly shift the results of the survey.

2. *Experiment #2: Ambiguous Bonus Agreement*

Experiment #1 dealt with a classic case of ambiguous language. In Experiment #2, we wanted to see how respondents decide a case that judges viewed as unambiguous, but where different judges assigned different (unambiguous) meanings to the contract language. We hoped that this exercise would begin to give us some clue as to how judges and respondents differ in their interpretation instincts.

Some of the classic cases about contractual ambiguity arise in employment bonus agreements.¹³¹ Such was the issue in an Illinois Supreme Court case, *Storybook Homes, Inc. v. Carlson*,¹³² in which the employer and its two employees agreed on a profit sharing bonus that stated:

[T]he bonus shall be computed in the following manner
0 to \$10,000 Net Profit No Bonus
\$10,000 to \$20,000 a maximum bonus of 5% Shall be paid to both
[employees]
\$20,000 and over a Maximum bonus of 22% shall be paid to each.¹³³

¹³⁰ The results are highly statistically significant. It is also assuring that the control group's behavior mimicked that of the Wave 1 respondents: Almost the same ratio of respondents sided with the insurer (47% to 46%).

¹³¹ See, e.g., *Rodriguez v. Miranda*, 507 S.E.2d 789 (Ga. Ct. App. 1998) (determining whether a terminated employee was entitled to a prorated share of the yearly bonus set forth in his employment contract); *Varney v. Ditmars*, 111 N.E. 822 (N.Y. 1916) (involving a contract term in which the employee was promised a "fair share" of the profits); *Salvaggio v. New Breed Transfer Corp.*, 564 S.E.2d 641, 643 (N.C. Ct. App. 2002) (determining whether voluntary termination by an employee precludes collecting a bonus for the time they worked).

¹³² 312 N.E.2d 27 (Ill. App. Ct. 1974).

¹³³ *Id.* at 28.

Subsequently, the firm earned profits exceeding \$20,000. The question came up: Is the employee entitled to 22% of all profits (employee's position) or only 22% of those profits above \$20,000, and a smaller percentage (0% and 5%) of the rest (employer's position)? The question was submitted to the jury, which sided with the employee. The trial judge then entered judgment N.O.V. in favor of the employer. The judge found that the bonus term was unambiguous because only the employer's interpretation of the language made business sense. The judge also invoked a somewhat obscure canon of interpretation—the last antecedent clause canon, which applies the 22% provision only to the last clause in the excerpt—to support his conclusion.¹³⁴ The Illinois Appellate Court affirmed the trial court's judgment unanimously on both grounds, saying that the language at issue “so overwhelmingly” favored employer's interpretation that “no contrary verdict . . . could ever stand.”¹³⁵

We puzzled over this case, because in our view the express term is easily susceptible to both interpretations—a fact that doctrinally ought to have left the resolution in the hands of the jury. We presented the survey respondents with a simplified version of the bonus clause, which evoked the same interpretation challenge:

The employee will receive an annual bonus in the following manner:

- \$1 to \$20,000 store profits – 5% Bonus;
- \$20,000 store profits and over – 20% Bonus.

Respondents were told that the firm earned a profit of \$25,000 and were asked to pick between the employee's claim of a \$5000 bonus (20% of the entire profit) and the employer's interpretation of a \$2000 bonus (5% of the first \$20,000 profit plus 20% of the profit above the \$20,000 threshold). We explained to the respondents how each of the contending bonus figures was computed. The distribution of results from the question “How much does the employer owe the employee?” was:

¹³⁴ *Id.* The trial judge held that under the employee's interpretation a few dollars difference could have major economic consequences. *Id.* at 30. It was also unlikely that the firm intended to award each employee (even two important employees) 22% of total profits. *See id.* at 30 (noting the “exceedingly large percentage of profit consumed by such an interpretation”).

¹³⁵ *Id.* at 29–30 (quoting *Pedrick v. Peoria & E. R.R. Co.*, 229 N.E.2d 504, 511 (Ill. 1967)).

TABLE 2A. BONUS AGREEMENT

<i>Answers</i>	<i>Frequency</i>	<i>Percent</i>
DEFINITELY \$5000	264	26.4
PROBABLY \$5000	207	20.7
It is COMPLETELY UNCERTAIN whether it is appropriate to pay the employee \$5000 or \$2000	248	24.8
PROBABLY \$2000	166	16.6
DEFINITELY \$2000	115	11.5
Total	1000	100.0

Interestingly, the respondents had a reaction similar to the jury's reaction—the employee's (more remunerative) interpretation of the contract language was the more reasonable interpretation. About 47% of respondents in Wave 1 thought the employee's interpretation was definitely or probably correct, versus 28% who thought the employer's interpretation was definitely or probably correct. Among those who took a position, almost two-thirds (63%) sided with the employee's interpretation.¹³⁶ Many respondents answered that the language was completely ambiguous, perhaps because the question required arithmetic computation exceeding their level of numeracy.¹³⁷

What was going on? Possibly, the term is equally susceptible to two legitimate interpretations, yet we obtained one-sided results because respondents favor the “little guy.” This could be thought of as a bias, which would be problematic (even though it might afflict juries as well). Or, it could be thought of as an intuitive manifestation of the *contra proferentem* logic—a bias against the drafting party whose carelessness created the ambiguity.¹³⁸ To arbitrate between the two conjectures, we reran the survey in Wave 2. This time, we showed half of the respondents the same language as in Wave 1 (the control group), and the other half a clearer text:

¹³⁶ This ratio—only 34% of those who did not choose “uncertain” favored the employee's interpretation—was exactly replicated in Wave 2. A statistical test confirmed that this ratio is highly statistically significant as a measure of the actual ratio in the population.

¹³⁷ See MARK KUTNER ET AL., NAT'L CTR. FOR EDUC. STATISTICS, U.S. DEP'T OF EDUC., *LITERACY IN EVERYDAY LIFE: RESULTS FROM THE 2003 NATIONAL ASSESSMENT OF ADULT LITERACY* 4, 13 (2007) (reporting that 22% of the adult population is at “below basic” level of numeracy—able to only locate numbers and perform simple operations (primarily addition) with very concrete and familiar information).

¹³⁸ *Sturm v. United States*, 421 F.2d 723, 727 (Ct. Cl. 1970) (stating that the rule “puts the risk of ambiguity, lack of clarity, and absence of proper warning on the drafting party which could have forestalled the controversy; it pushes the drafters toward improving contractual forms; and it saves contractors from hidden traps not of their own making”).

The employee will receive an annual bonus of 5% on any store profits earned between \$1 and \$20,000. If the store earns profits above \$20,000, the employee will receive a 20% bonus only on those amounts above \$20,000.

This new language very clearly favors the employer's position. Indeed, it would be challenging to draft a clearer but still succinct clause. The results were the following:

TABLE 2B. REDUCED AMBIGUITY BONUS AGREEMENT

Answers	Original Bonus Agreement (Control Group)		Clearer (Pro-Employer) Bonus Agreement (Treatment Group)	
	Frequency	Percent	Frequency	Percent
DEFINITELY \$5000	181	28.5	101	15.4
PROBABLY \$5000	139	21.9	109	16.6
It is COMPLETELY UNCERTAIN whether it is appropriate to pay the employee \$5000 or \$2000	134	21.1	116	17.7
PROBABLY \$2000	107	16.9	103	15.7
DEFINITELY \$2000	74	11.7	227	34.6
Total	635	100.0	656	100.0

The control group, which saw the same language as in Wave 1, behaved exactly as the Wave 1 respondents, with a ratio of 50% to 29% favoring the employee, and 21% saying the contract's meaning was completely uncertain. But the treatment group, which saw the revised language that clearly favors the employer, flipped, with a ratio of 50% to 32% now favoring the employer.¹³⁹ The percentage of respondents saying the employer's position was definitely correct tripled. It is admittedly disquieting that even such clear language did not elicit a more one sided distribution of results, suggesting the existence of some pro-little guy sentiment, or perhaps some numeracy deficiencies on the part of respondents. But the strong, statistically significant majority should suffice to point the courts toward the pro-employer interpretation.¹⁴⁰

¹³⁹ The results are statistically significant, at the 99% level.

¹⁴⁰ An alternative approach in a case like *Storybook* would be to use the survey evidence to calculate a weighted average, moving the law away from a binary choice. When 64% of (non-"uncertain") respondents favor the employer and 36% favor the employee, the bonus would amount to \$3080 (calculated as the weighted average $0.64 \times \$2000 + 0.36 \times \5000). The weighted average approach would also permit researchers to utilize some of the granularity that arises from a five-point scale rather than a two-point or three-point scale, by giving greater weight to "definitely" responses.

In sum, Experiment #2 provides ample reason to critique what the judges did in *Storybook*. The state supreme court's determination that the bonus provision "overwhelmingly favored" *Storybook*'s interpretation was probably not an accurate reflection of how the employees understood the contractual language. Only 28% of our sample agreed with the court's characterization. The jury was on to something important that the judges completely missed. Moreover, as our Wave 2 data shows, the employer-drafter could have made small tweaks to its employment agreement that would have caused the bonus term to clearly convey the drafter's intent.

3. *Experiment #3: Interpretation Against the Drafter*

What majority should suffice for a court to hold that a particular interpretation prevails? How should the court account for the frequency of the "uncertain" response? And do these heuristics depend on who among the parties drafted the contract? To gain some insight on these issues, we turned again to insurance contracts.

Experiment #3 was based in part on the facts of *Vargas v. Insurance Co. of North America*, a Second Circuit opinion interpreting the language of an aviation insurance policy.¹⁴¹ The policy in question applied "only to occurrences, accidents or losses which happen . . . within the United States of America, its territories or possessions, Canada or Mexico."¹⁴² The crash of a small airplane occurred as the plane was flying from New York to Puerto Rico, in international waters twenty-five miles west of Puerto Rico. The pilot policyholder's estate argued that a trip from New York to the U.S. territory of Puerto Rico meant that the crash was covered. The insurer argued that because the fatal crash occurred outside of Puerto Rico's territorial waters, the policy did not apply.

The trial court granted summary judgment for the insurer, but the Second Circuit reversed, finding the language to be ambiguous and invoking the rule that ambiguous language (which could have been made clearer by better drafting) should be interpreted against the drafter.¹⁴³ We showed respondents in Wave 1 a vignette based on *Vargas*, and obtained these results:

¹⁴¹ 651 F.2d 838 (2d Cir. 1981).

¹⁴² *Id.* at 839.

¹⁴³ *Id.* at 838, 840–42 ("The policy in this case is . . . ambiguous . . . and the insurer bears the responsibility for not adopting a clear exclusion.")

TABLE 3A. PUERTO RICO AIRPLANE CRASH INSURANCE

<i>Answers</i>	<i>Frequency</i>	<i>Percent</i>
Crash DEFINITELY covered	125	12.5
Crash PROBABLY covered	134	13.4
It is COMPLETELY UNCERTAIN whether the crash is covered	226	22.6
Crash PROBABLY NOT covered	308	30.8
Crash DEFINITELY NOT covered	207	20.7
Total	1000	100.0

A strong majority favored the insurer's interpretation over the insured's (51% to 26%), consistent with the trial court's position. The result is statistically significant, which means that we can be confident that more than 50% of the population would favor the insurer's interpretation. This is a strong margin of prevalence, where two-thirds of respondents who chose one of the two positions sided with the insurer. We think that such a margin should suffice to defeat the *contra proferentem* presumption.

This is not to say that the insurer drafted the clearest of all languages. The insurer could have removed some of the ambiguity with better drafting, and we now show that if parties draft more carefully, they can win interpretation battles with more convincing majorities. In Wave 2, we chose a contested aviation insurance term from a different case, *Security Insurance Co. of Hartford v. Andersen*.¹⁴⁴ In that case, the policy contained a condition for coverage requiring that the pilot hold a valid and current medical certificate attesting fitness to fly.¹⁴⁵ More precisely, it read, "this Policy shall apply only while the aircraft is operated in flight by the pilot(s) designated below and then only if the said pilot . . . holds a valid and current medical certificate of the appropriate class[.]"¹⁴⁶

At the time of the crash, the pilot did not have such a certificate, because it had expired a few months before his fatal accident. The policyholder's estate argued that since there was no causal connection between the failure to obtain the medical certification and the crash, the policy should cover the crash. The intermediate appellate court accepted this argument, as some other jurisdictions had done, and determined that the crash would be covered in the absence of a causal link.¹⁴⁷ The Supreme Court of Arizona unanimously disagreed,

¹⁴⁴ 763 P.2d 246 (Ariz. 1988).

¹⁴⁵ *Id.* at 247–48.

¹⁴⁶ *Id.* at 248 (emphasis omitted).

¹⁴⁷ *Id.* at 249.

finding that the insurance policy language at issue was “completely unambiguous.”¹⁴⁸ Under the language of the agreement, there was no need for the insurance company to show any causal connection between the absence of a medical certification and the cause of the crash.¹⁴⁹

We showed respondents a vignette based on *Andersen* in Wave 2 of the survey. This time, respondents overwhelmingly agreed with the insurer’s interpretation (65%, versus 22% favoring the policyholder, and only a relatively low 13% of the sample choosing “completely uncertain”).¹⁵⁰ Now, 75% of those siding with one of the two interpretations sided with the insurer. This 65-22 margin of prevalence is statistically significantly greater than the margin of prevalence (51-26) that we saw in the *Vargas* vignette. This level of consensus in *Andersen* would surely be adequate to overcome contra proferentem, resulting in a win for the insurer.

TABLE 3B. ARIZONA AIRPLANE CRASH INSURANCE

<i>Answers</i>	<i>Frequency</i>	<i>Percent</i>
Crash DEFINITELY covered	146	11.3
Crash PROBABLY covered	137	10.6
It is COMPLETELY UNCERTAIN whether the crash is covered	168	13.0
Crash PROBABLY NOT covered	353	27.3
Crash DEFINITELY NOT covered	489	37.8
Total	1293	100.0

Under any workable legal regime, the answer provided by the Arizona Supreme Court in *Andersen* has to be the correct one, given the clarity of the contractual provision. At the same time, even when confronted with language that the Arizona Supreme Court held to be “completely unambiguous,” 22% of the sample reached the opposite interpretation. We therefore think that the *Andersen* case can help “normalize” the results of interpretation surveys by establishing benchmarks for preponderance. The standard for ambiguity when

¹⁴⁸ *Id.*

¹⁴⁹ *Id.* at 250–51.

¹⁵⁰ Wave 2 respondents were randomly assigned to two different conditions reflecting legally irrelevant alterations to the vignette. The control group was told (correctly) that the policy required both a valid medical certificate and an FAA certification to fly. The treatment group saw that the policy required only a valid medical certificate. The respondents were given no information to suggest that the presence of an FAA certification was contested by the insurance company, so the experimental manipulation merely lengthened the vignette and added superfluous complexity. The change made no difference, as there were no significant differences across conditions.

using the consumer survey method cannot be whether 15% or 20% of a sampled population embraces the minority view. Nor should the fact that 15% of a sample says language is completely ambiguous cause a court to agree with them. Some respondents may be answering at random, misunderstanding the language and survey questions, or focusing entirely on the equities of the vignettes, suggesting that even in clear-cut cases a sizeable minority of respondents will go the other way.

While further work is needed to generate effective thresholds, we think it can be done. If surveys of clear and unambiguous language generate, for example, an average “uncertain” response rate of 20%, this should be the new “zero” and only rates exceeding this baseline should count as true votes for ambiguity. Likewise, if a clear term garners only an average support of 75% of those siding with one of the two interpretations, this should be the new “100%” and majority rates should be normalized in relation to it.¹⁵¹ Perhaps there should also be a sliding scale: the larger the “uncertain” block, the greater the majority required among the rest.

4. *Experiment #4: A “Little Guy” Effect?*

As we hinted in our discussion of Experiments 2 and 3, some respondents to surveys more often share the experiences and perspectives of the “little guy.” As a result, aggregate responses might be biased against corporations and other large entities. While the two previous experiments gave us confidence that most respondents can be trusted to side even with employers or the insurance industry when the language is clear, we nevertheless wanted to see whether there is a “little guy” effect in a context in which it is most likely to arise. The context we chose is a covenant not to compete in an employment contract. Such covenants restrict the ability of workers to switch jobs and are regarded by many as unfair. They are illegal or borderline illegal in some jurisdictions, and we think there is a widespread sentiment to limit their application, especially when the employee is fired.¹⁵²

¹⁵¹ For example, under these assumptions, survey results of an unambiguous pro-X term would be, on average, 60% to 20% in favor of X among those voting for one of the two interpretations, with 20% choosing “uncertain.” The case for X should not be weakened by its failure to exceed the 60% threshold or by the presence of 40% who did not favor it.

¹⁵² Courts generally only enforce these covenants if their time and geographic restrictions are reasonable. *See, e.g., Meyer v. Wineburgh*, 110 F. Supp. 957, 959 (D.D.C. 1953) (holding that a restrictive covenant is valid if it is “reasonably limited as to time and territory” and does not “work an unfair hardship upon the restricted party”); *see also In re UFG Int’l, Inc.*, 225 B.R. 51, 56 (S.D.N.Y. 1998) (“The cases which hold that a covenant not to compete is unenforceable against an employee who is terminated without cause are

In *Cambridge Engineering v. Mercury Partners*, an employee of company A who worked as a sales representative signed a covenant not to compete.¹⁵³ After he was fired, he took a job with company B, a competitor of company A. The non-compete agreement at issue established in relevant parts that: “Employee shall not, for a period of 24 months following the termination of his/her employment, . . . engage in any activity for or on behalf of Employer’s competitors, or engage in any business that competes with Employer, anywhere in the United States or Canada.”¹⁵⁴

Does this language preclude an employee of company A from working for company B? The first possible interpretation is broad: Any job with company B is prohibited, even if it does not directly compete with company A, because company B is a competitor. Under this interpretation, it would be prohibited for the employee to take a job even as a janitor for company B. The second possible interpretation is narrow: The prohibition applies only to jobs with company B that directly compete with company A.¹⁵⁵ The court chose the broad interpretation, relying on an interpretive canon—the so-called rule against “surplusage.”¹⁵⁶

Survey respondents were presented with a vignette modeled on the *Cambridge* case. They were told that the employee worked as a salesperson for company A and agreed to a non-compete clause that said:

Employee may not engage in any activity for competitors of [company A], or engage in any business that competes with [company A].

Respondents were told that the employee took a job as a human resources officer with company B, thus not directly competing with company A’s sales staff. They were asked whether the employee was prohibited by the non-compete clause from taking this job with com-

premised on the unfairness of permitting an employer who has destroyed the mutuality of obligation on which the covenant rests to benefit from the covenant.”).

¹⁵³ 879 N.E.2d 512 (Ill. App. Ct. 2007).

¹⁵⁴ *Id.* at 517.

¹⁵⁵ Ironically, in *Cambridge* it was the employee who argued for the broad interpretation. *Id.* at 524–25. Under Illinois law, a non-compete that is overbroad in scope—that forbids the employee from taking new positions that do not directly compete with the original employer—is against public policy and entirely unenforceable. *See id.* at 523 (explaining that in order for the covenant to be enforceable, its territorial scope must be reasonable and coextensive with the area in which the employer does business). Conversely, it was the company that advocated for the narrow interpretation, hoping that it would survive the enforceability test. *Id.* at 524–25.

¹⁵⁶ *Id.* at 525. The court emphasized that under the narrow interpretation, the language “for or on behalf of Employer’s competitors” would be rendered superfluous. *Id.*

pany B. Identical language was tested in both waves of the survey. Here are the results:

TABLE 4. BONUS AGREEMENT

<i>Answers</i>	<i>Wave 1 Frequency</i>	<i>Wave 1 Percent</i>	<i>Wave 2 Frequency</i>	<i>Wave 2 Percent</i>
DEFINITELY prohibited	307	30.7	501	38.7
PROBABLY prohibited	301	30.1	295	22.8
COMPLETELY UNCERTAIN	239	23.9	241	18.6
PROBABLY NOT prohibited	97	9.7	144	11.1
DEFINITELY NOT prohibited	57	5.7	111	8.6
Total	1001	100.0	1292	100.0

Wave 1 respondents agreed lopsidedly (and in a statistically significant manner) with the broad interpretation—that the non-compete agreement prohibited the employee from taking any position with company A’s competitor (61% versus 15%). These results were replicated in Wave 2 of the survey. The supermajoritarian interpretation in both waves is opposed to the superficially apparent interests of the “little guy.”¹⁵⁷ It is encouraging that the results did not vary materially over time, suggesting that a firm could safely field test contract terms at early phases and accurately anticipate their subsequent legal exposure.

5. *Experiment #5: Interpretation of Consumer Standard Form Contract*

Our final illustration concerns the interpretation of a term in Gmail’s privacy policy.¹⁵⁸ We included it here, even though these findings have been reported elsewhere, because it illustrates additional promise for the survey interpretation method. The case itself is still pending, but it has dragged on for years and proved costly to resolve.

¹⁵⁷ An alternative possibility is that some respondents harbored antipathy towards an employee seen as disloyal, rather than having pro-little guy sentiment. These varied moral frames would counteract one another in our data. It is also possible that some tiny share of the well-informed respondents who understood the counterintuitive effect—that a broad interpretation would actually benefit the employee because it renders the entire non-compete covenant void—voted for the broad interpretation strategically. If such an effect existed, it would be small, and it would bias results against the result reached by the court, which was that the covenant as written precluded work of any type for a competitor.

¹⁵⁸ This illustration comes from a separate experimental survey, conducted prior to the other surveys reported above and published separately. See Lior Jacob Strahilevitz & Matthew B. Kugler, *Is Privacy Policy Language Irrelevant to Consumers?*, 45 J. LEGAL STUD. 669, 677–80 (2016) (describing the results of the Gmail experiment).

The case displays some of the problems with the existing court-based interpretation. Could the survey method do better?

In the Gmail litigation, Google was sued for scanning users' email messages, a practice that—unless agreed upon by users—would be regarded as a violation of federal wiretap laws and subject to enormous damages.¹⁵⁹ In its defense, Google invoked a term in its privacy policy, by which users allegedly grant consent to scan their email messages. The parties contested the meaning of the term and whether it indeed granted Google explicit consent from users. The early-stage litigation turned on the interpretation of one paragraph.¹⁶⁰ In denying Google's motion to dismiss, the federal district court held that the term in the privacy policy was ambiguous.¹⁶¹ It was therefore a triable issue whether the term was effective to secure users' consent to the company's practice.

As it happens, Yahoo was also sued for having engaged in email scanning conduct very similar to Google's conduct, and the case was assigned to the same federal judge.¹⁶² Here, the judge found that Yahoo's privacy policy was unambiguous and provided Yahoo users with notice adequate to secure their consent.¹⁶³ In short, the court interpreted the two contracts differently: Only Yahoo users, but not Gmail users, received unambiguous notice and consented to the content analysis.

When randomly assigned respondents read either the legally unambiguous Yahoo policy or the legally ambiguous Gmail policy, there was no statistically significant difference in the percentage of readers who regarded the policies as sufficiently informative to obtain their consent to the practices at issue. Asked if the contract term allowed Gmail to scan their emails, a surprisingly strong majority of respondents said that it was in fact unambiguous in informing consumers about the scanning of their emails.¹⁶⁴ This was the case even though respondents regarded the email provider's conduct at issue (automated content analysis of user emails for the purposes of showing user personalized ads) as quite intrusive.

¹⁵⁹ *In re Google Inc. Gmail Litig.*, No. 13-MD-02430, 2013 WL 5423918, at *1 (N.D. Cal. Sept. 26, 2013).

¹⁶⁰ In reality, there were three separate versions of this term, as Google has changed the text over the years. *Id.* at *13–14.

¹⁶¹ *Id.* at *14.

¹⁶² *In re Yahoo Mail Litig.*, 7 F. Supp. 3d 1016 (N.D. Cal. 2014).

¹⁶³ *Id.* at 1029.

¹⁶⁴ Strahilevitz & Kugler, *supra* note 158, at S77–78. The experiment differed from the experiments for this article in one aspect: Respondents did not have a choice of “completely ambiguous.” *Id.* at S78.

TABLE 5. GMAIL/YAHOO PRIVACY AGREEMENT

<i>Answers</i>	<i>Percent</i>
DEFINITELY ALLOWED	26.6
PROBABLY ALLOWED	38.5
PROBABLY NOT ALLOWED	13.4
DEFINITELY NOT ALLOWED	21.3
Total	99.8

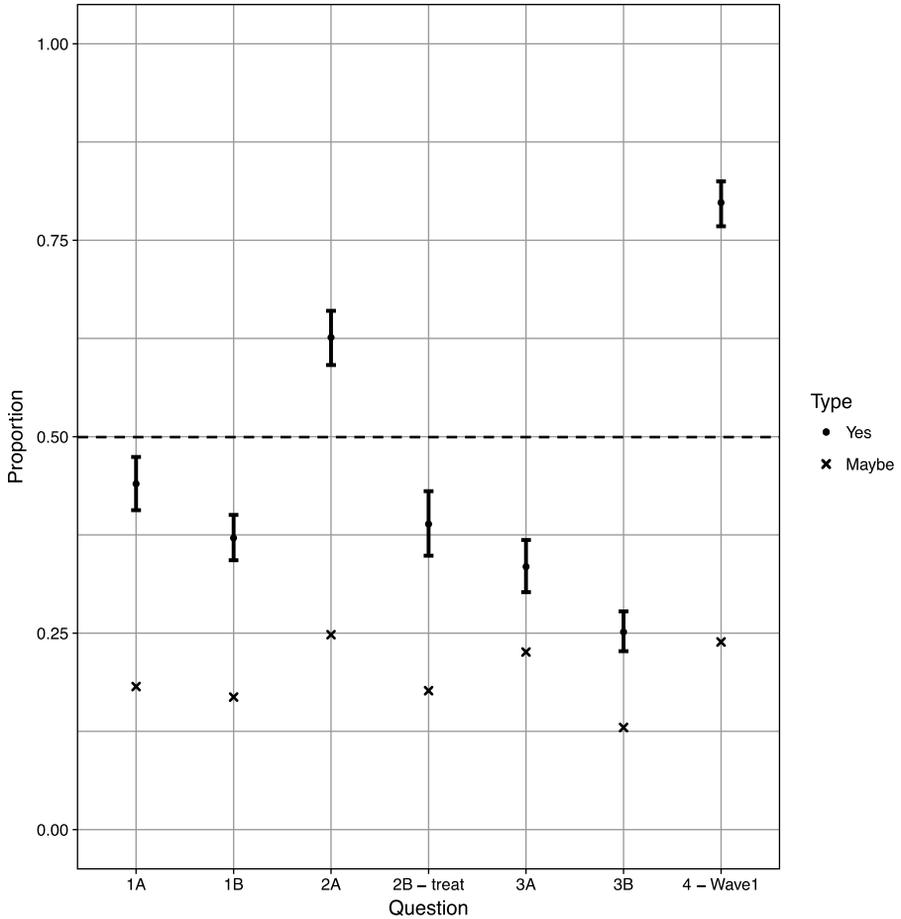
This is disconcerting. Changes in the language that a court deemed legally decisive were regarded as irrelevant by participants in a randomized experiment.¹⁶⁵ The *Google* and *Yahoo* cases therefore represent further instances in which courts, adjudicating lengthy and costly interpretation battles, are characterizing the common understanding of contractual language in ways that are at odds with the actual reactions of a representative sample of lay Americans.

C. Summary

The results of experiments one through four can be collected in the following figure:

¹⁶⁵ Strahilevitz and Kugler obtained similar results in a separate study of Facebook users that presented them with precise and vague policies describing Facebook's use of facial recognition software to facilitate photo tagging. *See id.* at S80–83 (showing no significant difference in respondents' interpretation despite changes in the wording of the policy language). The Facebook litigation is still pending at the trial court level, so there has been no judicial resolution of the interpretive issues in that case.

FIGURE 1. SUMMARY RESULTS



Each dot represents the proportion of respondents who chose the first interpretation. The statistical tests we conducted are represented by the 95% confidence interval surrounding these point estimates. It is striking that none of the survey results show respondents dividing evenly between the interpretations. (This can be seen by the fact that none of the confidence intervals cross the 0.5 line.) In addition, the X dots represent the ratio of respondents who chose “uncertain,” which varied from 12.5% to 25%.

Each of the cases that inspired the surveys and experiments above was a tough interpretive nut for courts to crack. Each case proved hard enough to warrant appellate litigation that involved reversals of lower court interpretations. Some judges drew on lay intuitions that were not backed by fully developed evidential records.

Others drew on interpretive canons that have little to do with how most people read texts.

Sometimes, the judiciary's judgments were entirely consistent with the respondents' majority. This was the result in *Moore*'s dissent as well as in *Andersen*, *Cambridge*, and *Yahoo* (where courts and respondents agreed the language was unambiguous). But the *Moore* majority, *Vargas*, *Storybook*, and *Google* missed the mark compared to the majority of respondents' understanding. The *Storybook* bonus case is the most striking, because the court embraced an unambiguous meaning of words that was opposite to the jury's interpretation,¹⁶⁶ and we found that a resounding majority among our respondents agreed with the jury. The *Google* court held that privacy policy language was too ambiguous to secure user consent to content analysis of their emails,¹⁶⁷ but respondents who read the policy language overwhelmingly disagreed. We could see the clash between the court and the survey respondents in *Storybook* coming, but the *Moore* and the *Google* results took us by surprise, indicating that shifting to the consumer survey methodology would not only change the process of interpreting contracts but would at least occasionally change the substantive interpretations of contracts as well.

If the heterogeneity of the response data in our sample shows us anything, it is that individual judgments and responses can be quirky and mystifying, but majoritarian judgments about contractual meaning are comprehensible. However anecdotal, our surveys begin to demonstrate that respondents are good at identifying ambiguity when it clearly exists, and that they shift in the right direction when the language is made clearer through experimental manipulations. Importantly, they are not inevitably biased against business parties. Further research is necessary to identify with precision how surveys differ systematically from court interpretation. But our purpose in this series of experiments was to demonstrate the plausibility and practicality of the survey interpretation method.

As we turn to Part IV, in which we discuss the normative justifications for the survey interpretation method, the contrasts between judicial and lay interpretations raise a profound question. Is it preferable to base the binding meaning of consumer contracts on the majoritarian understanding of those texts rather than judicial interpretations? Are the intuitions of like transactors a better yardstick than the professional judgments of courts that are based on prior prece-

¹⁶⁶ *Storybook Homes, Inc. v. Carlson*, 312 N.E.2d 27, 29–30 (Ill. App. Ct. 1974).

¹⁶⁷ *In re Google Inc. Gmail Litig.*, No. 13-MD-02430, 2013 WL 5423918, at *14 (N.D. Cal. Sept. 26, 2013).

dents, policy considerations, interpretive canons, or judicial assessments of parties' purposes?

IV

THE CASE FOR SURVEY INTERPRETATION

Parts II and III presented a novel approach to contract interpretation that would potentially overcome the problems that Part I identified with existing doctrine. The focus of Parts II and III was descriptive. They intended to answer the question whether the survey interpretation method is *feasible*. To that end, Part II described how the new system would work and the wide reception and success it had in the neighboring areas of trademark and unfair competition law. And Part III reported the findings of a pilot experiment, in which we implemented the proposed regime in a variety of old interpretation challenges and obtained promising results.

We now turn to the question that is perhaps the most fundamental: Why survey interpretation? It is not enough to show that the method is feasible, nor (as we will argue later) that it avoids doctrinal knots. The survey interpretation method is a significant change to the *practice* of contract interpretation, and it needs to be justified.

In this Part, we present a normative defense of the survey interpretation method, based on several arguments. First, interpretations based on surveys further the goal of giving a text the understanding assigned by those for whom the text was written. Second, survey-based interpretation in many—or perhaps even most—cases will be cheaper than existing interpretive approaches. Third, survey-based interpretation would render the interpretation of contracts more predictable. Fourth, it would displace “normative interpretation.” And fifth, the survey interpretation method may have beneficial dynamic effects, resulting in the simplification of contract terms.

A. *Enforce the Terms People Expect*

The core of the normative case for using the survey interpretation method is straightforward: Contracts should have the meaning that the parties to the transaction assign to the text.¹⁶⁸ It is pointless to ask the actual parties in the litigation what the text meant to them when they formed the contract, because they will bend their answers to fit their litigation goals. So the law should instead ask disinterested people just like them.

¹⁶⁸ See, e.g., *Opals on Ice Lingerie v. Bodylines, Inc.*, 320 F.3d 362, 372 (2d. Cir. 2003) (holding that a “meeting of the minds” is the fundamental basis of a valid, enforceable contract).

The case for the survey interpretation method is pronounced when this method yields a meaning that differs from the existing method, with its reliance on legally trained interpretation. A consumer, employment, or insurance contract is written for lay people, not experts, and the meaning of its provisions should be determined by its lay, not sophisticated, understanding. A commercial contract is written for merchants, and its meaning should be determined by surveying people in the trade, harnessing their industry-specific literacy. And a sophisticated drafted-from-scratch contract is written by and for the attorneys on each side, and thus other like professionals should be polled to determine the meaning.

Let us focus on mass contracts that govern transactions with lay people for the moment. It might be thought that, unlike trademarks and advertising messages, standard contract terms are not written for the lay person. Surely, most people do not pay attention to boilerplate; why should they be surveyed as to its meaning? The primary answer (albeit disappointingly simple) is: It's the law! This criterion—how an ordinary recipient of a contractual message would understand it—is a touchstone of contract law, used to determine the meaning of precontractual representations, offers, and contractual terms. “[W]e give words their ‘ordinary meaning,’ viewing the subject of the contract ‘as the mass of mankind would view it.’”¹⁶⁹

Courts routinely recognize that they ought to uncover the meaning that lay parties would assign to the language, working to put themselves in the shoes of the non-legally trained.¹⁷⁰ But can judges truly set aside the influences of their prior knowledge and expertise and imagine what a term means to non-experts? As one court observed, trial judges—with background, knowledge, and experience unlike that of most consumers—are hardly in a position to understand consumer-facing communications the same way that consumers do.¹⁷¹ The survey method would help courts overcome such cognitive and information biases.

Compare this crowdsourcing method of interpretation to the existing legalistic alternative, which relies on technical canons of interpretation. The use of some of these archaic techniques has fortunately subsided over time, but they are still a central feature of court-centered interpretation. Some canons are intuitive and might be fol-

¹⁶⁹ *All-Ways Logistics, Inc. v. USA Truck, Inc.* 583 F.3d 511, 516 (8th Cir. 2009) (quoting *Coleman v. Regions Bank*, 216 S.W.3d 569, 574 (Ark. 2005)).

¹⁷⁰ *See, e.g., Bristol West Ins. Co. v. Wawanesa Mut. Ins. Co.*, 570 F.3d 461, 464 (1st Cir. 2009); *Sargent Constr. Co. v. State Auto Ins. Co.*, 23 F.3d 1324, 1327 (8th Cir. 1994).

¹⁷¹ *Verizon Directories Corp. v. Yellow Book USA, Inc.*, 309 F. Supp. 2d 401, 407 (E.D.N.Y. 2004).

lowed by survey respondents. Others, like the last antecedent clause rule (relative and qualifying words or phrases are to be applied only to the words or phrases that immediately precede them),¹⁷² *ejusdem generis* (the meaning of the word is interpreted by the company it keeps),¹⁷³ or *expressio unius est exclusio alterus* (to express or include one thing implies the exclusion of another),¹⁷⁴ are more technical and might not resonate at all with the ways lay people interpret texts. Indeed, there is an accumulating body of evidence to suggest that interpretive canons are foreign not only to naïve laypeople but to sophisticates as well. Capitol Hill staffers who draft laws are often unfamiliar with the interpretive canons courts use to assess their handiwork.¹⁷⁵ More troublingly, recent qualitative work suggests that courts interpreting choice of law clauses in contracts often apply canons that even the lawyers who negotiate these clauses do not expect them to use, resulting in interpretations that are at odds with what scores of lawyers said their clients intended.¹⁷⁶

Even in those rare instances where a contract interpretation case goes to a jury—which is existing law’s best attempt to uncover the meaning that the “mass of mankind” gives to the contract—the outcome is inferior to the survey method. First, the larger “*N*” in the survey reduces the variance and enhances the validity of the sample as a representation of the population at large. More subtly, because jurors get to deliberate, they are not representative and are not independent “draws.” Consumers reading a contract do not deliberate, but rather follow a more expeditious thought process, which is better simulated by survey respondents. If they ever read the text, consumers do it alone, not with eleven other people in the room. Group deliberations, as conducted by juries, are known to lead

¹⁷² See, e.g., *Stewman Ranch, Inc. v. Double M. Ranch, Ltd.*, 192 S.W.3d 808, 812 (Tex. App. 2006).

¹⁷³ See, e.g., *Aspen Advisors LLC v. United Artists Theatre Co.*, 861 A.2d 1251, 1265 (Del. 2004); 242-44 E. 77th St. Greater N.Y. Mut. Ins. Co., 815 N.Y.S.2d 507, 510 (App. Div. 2006) (explaining that when a general provision follows a series of specific provisions, the general provision is interpreted as being of the same type or class as the specific provisions that preceded it).

¹⁷⁴ See, e.g., *Mastrocovo v. Capizzi*, 930 N.Y.S.2d 141, 142 (App. Div. 2011).

¹⁷⁵ 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, 926–29 (2013).

¹⁷⁶ Coyle, *supra* note 95, at 696–701.

towards polarization,¹⁷⁷ often according some jurors disproportionate influence.¹⁷⁸ Surveys dodge this problem.

The survey interpretation method cannot ensure that every party to a contract gets the deal that he or she expected, had they read the terms. People are too heterogeneous and problems of proof are too significant for any reform to achieve that goal. Yet the method can ensure that most parties to an agreement are receiving the terms that they expected or that they would have expected had they taken the time to read and digest the terms of the agreement. That strikes us as a significant benefit regardless of whether it is couched in deontological terms that emphasize individual autonomy or consequentialist terms that emphasize the value of Pareto-improving bargains.

B. Costs of Interpretation

Interpretation of contractual text may be the most frequent ground for dispute in contract law.¹⁷⁹ These disputes are costly, in part because they require courts to determine what extrinsic evidence to consult (and then to consult it), but primarily because they drain expensive lawyer time.¹⁸⁰

Survey interpretation is becoming increasingly cheap. In the past, a major criticism of surveys in trademark litigation was their cost.¹⁸¹

¹⁷⁷ See Doron Teichman & Eyal Zamir, *Judicial Decision-Making: A Behavioral Perspective*, in THE OXFORD HANDBOOK OF BEHAVIORAL ECONOMICS AND THE LAW 664, 677 (Doron Teichman & Eyal Zamir eds., 2014). See generally Cass R. Sunstein, *Deliberative Trouble? Why Groups Go to Extremes*, 110 YALE L.J. 71 (2000) (investigating the phenomenon of group polarization).

¹⁷⁸ See, e.g., Erin York Cornwall & Valerie P. Hans, *Representation Through Participation: A Multilevel Analysis of Jury Deliberations*, 45 LAW & SOC'Y REV. 667 (2011) (confirming that social status affects participation in jury deliberations); Nancy S. Marder, Note, *Gender Dynamics and Jury Deliberations*, 96 YALE L.J. 593, 594–98 (1987) (finding that women generally have lower participation rates in jury deliberations than men).

¹⁷⁹ See generally Jeffrey M. Lipshaw, *Metaphors, Models, and Meaning in Contract Law*, 116 PA. ST. L. REV. 987, 989 (2012) (“[I]t is a fair observation that only a tiny portion of the first-year contracts course involves the issue of contract interpretation. Yet practitioners know that the real world of contracts is almost exclusively about negotiating and writing documents and perhaps interpreting them later (whether or not they get litigated).”).

¹⁸⁰ For a discussion of the costs of contract interpretation and the disputes arising therefrom, see Richard A. Posner, *The Law and Economics of Contract Interpretation*, 83 TEX. L. REV. 1581, 1590–96, 1612–14 (2005).

¹⁸¹ See Diamond & Franklyn, *supra* note 44, at 2061 (arguing that cost might lead “[c]lients who may benefit from surveys” to be “potentially priced out of court”); see also Heymann, *supra* note 64 (“The effect of a survey in a trademark case is as much about which party has the resources to fully commit to the survey process as it is about a search for the truth about consumer perception.”); Thornburg, *supra* note 47, at 717 (“[C]urrent survey experts in California charge between \$450 to \$600 per hour and require support staff billing at rates ranging between \$200 to \$300 in orchestrating the actual surveys.”).

The International Trademark Association, for example, reported the cost of surveys to be in the \$25,000 to \$150,000 range.¹⁸² But the rise of online panel surveys promises a significant reduction in the cost of conducting surveys.¹⁸³ Online surveys are quicker, less expensive, and can reach a population of consumers that is demographically representative and nationally dispersed, avoiding the costs of annoying people who do not wish to be surveyed.¹⁸⁴ Reputable online panels provide sufficient compensation to respondents to ensure serious and meaningful responses.¹⁸⁵ Online surveys have the distinct advantage of allowing respondents to examine visual stimuli and texts with adequate time to ponder what they see.¹⁸⁶ The selection of the respondent panels is fully transparent, conducted by survey firms that stake their business reputations on doing it right. Perhaps most surprising is the finding that online surveys obtain substantive results that are statistically equivalent to those obtained under more traditional survey modes.¹⁸⁷

Once contract language is tested and its meaning validated via surveys, it could be replicated widely within an industry. As it stands, law firms and lawyers often use the same boilerplate language for multiple clients involved with similar deals. Sometimes the boilerplate's meaning has been adjudicated, but often it has not. When such language is adjudicated via ordinary methods, costly surprises sometimes ensue. Having pretested language and discerned its likely meaning for consumers, a law firm might advise multiple clients to use the same boilerplate. In that way, the already reasonable costs of conducting surveys could be distributed.

¹⁸² INT'L TRADEMARK ASS'N, *supra* note 76, at 5; *see also* Lisa Larrimore Ouellette, *The Google Shortcut to Trademark Law*, 102 CALIF. L. REV. 351, 361 n.53 (2014) (quoting a contemporary statement from a practitioner that trademark surveys typically cost \$75,000 to \$150,000 in a case).

¹⁸³ The online survey firms we used for the experiments reported in Part III charge \$3.25 per respondent, for a fifteen to twenty minute survey. Telephone interviews, by contrast, cost on average \$20 per respondent. *See* Karin Braunsberger et al., *A Comparison of Reliability Between Telephone and Web-Based Surveys*, 60 J. BUS. RES. 758, 763–64 (2007) (concluding that the cost of each web-based survey amounted to only 29% of the cost of each telephone survey).

¹⁸⁴ Hal Poret, *A Comparative Empirical Analysis of Online Versus Mail and Phone Methodologies for Trademark Surveys*, 100 TRADEMARK REP. 756, 757 (2010).

¹⁸⁵ Unlike the respondents contacted by phone or while shopping in malls, they are not being asked to donate their time.

¹⁸⁶ *See, e.g.*, *Muha v. Encore Receivable Mgmt., Inc.*, 558 F.3d 623, 626 (7th Cir. 2009) (“[A] telephone survey is not an ideal method of testing the understanding of a written statement, since inflection can alter meaning and some written statements are easier to understand when read than when heard.”).

¹⁸⁷ Poret, *supra* note 184, at 807.

Low survey costs make the method more attractive, but also easier to manipulate. As the costs of conducting online surveys decline, the danger increases that a party will try out multiple versions of surveys, cherry pick the most favorable results, and bury the less favorable ones. There are charlatans operating in the trademark litigation landscape “who can essentially create a survey to show any desired finding . . . by either creating a skewed line of questioning or numerical manipulation of already acquired data.”¹⁸⁸ In some instances, presenting survey respondents with longer excerpts from a contract may alter the results, and disputes will arise between the parties as to how much contractual text and context needs to be presented to the respondents. In that sense, battles over the survey interpretation methodology would inevitably replace some existing fights over the scope of extrinsic evidence that ought to be admitted. But the aggregate costs will be reduced—because the question is now simpler—the methodological question of how to best capture the ordinary lay meaning, rather than the normative question of what is a reasonable interpretation.

As we discussed in Part II above, the use of neutral experts chosen either by the court or in the contract also would help address the biased survey design problem.¹⁸⁹ One attractive aspect of the survey method under the adversarial system is that it promotes immediate replication, and experts whose findings consistently were outliers might suffer reputational damage. Beyond that, we would anticipate an accumulation over time of a kind of common law of survey interpretation, as we see in the area of trademark surveys.¹⁹⁰

Discovery rules may further alleviate the cherry-picking problem. Litigants’ efforts to hide disappointing survey results have to be foiled, and what better way to do it than to make various versions of contractual language that were actually tested fair game in civil discovery? We might worry about a chilling effect—the possibility that discovery might diminish the propensity of drafters to employ the method. But the risk that discovery poses can be overcome if drafters pre-test the language prior to the contract and use only versions that perform well in such testing.

¹⁸⁸ Thornburg, *supra* note 83, at 97; *see also* MCCARTHY, *supra* note 43, at § 32:179 (discussing protection of surveys from discovery); Diamond & Franklyn, *supra* note 44, at 2059 (finding evidence of attorneys who collected surveys and buried those with bad results).

¹⁸⁹ *See Scott v. City of New York*, 591 F. Supp. 2d 554, 560 (S.D.N.Y. 2008), for an example of the mediation approach in action. *See also* Welter, *supra* note 111, at 209 (explaining that survey results are more likely to be accepted by both parties if the court appointed a neutral expert to conduct the survey).

¹⁹⁰ *See supra* text accompanying notes 75–76.

Even low-cost online surveys may be too expensive for non-represented plaintiffs bringing small claims. Such disputes rarely involve contract interpretation battles because the contract underlying the dispute is usually in standard form and unambiguously favors the drafting party. In the rare case where the result of a small claim turns on the meaning of the boilerplate, we think the use of surveys should be exempted. But in general, consumer complaints are solved either via less formal settings (e.g., arbitration or online dispute resolution), or in a highly formal class action procedure. Long and costly interpretation fights take place in the latter, and the survey method would only make these battles cheaper.

C. *The End of “Normative” Interpretation*

Delegating interpretation of contracts to surveys would advance largely textual interpretations assigned by “the mass of mankind,” but it would at the same time oust interpretations based on non-textual approaches. Interpretation doctrine sometimes tries to do more than elicit the literal or plain meaning. Courts sometimes try to look for the purpose of the parties (under the premise that such common purpose exists).¹⁹¹ Or, if a term may be consistent with two reasonable meanings, courts use interpretation doctrine to assign liability for such confusion.¹⁹² Survey interpretation would thus kill the enterprise of “normative” interpretation—the practice through which courts infuse contracts with content that promotes social ends.¹⁹³

For example, survey interpretation would cripple courts’ ability to promote economic efficiency by choosing the meaning that minimizes transaction costs or maximizes the parties’ joint surplus.¹⁹⁴ We are not aware, however, of any evidence that judges currently try to interpret

¹⁹¹ See RESTATEMENT (SECOND) OF CONTRACTS § 202(1) (AM. LAW INST. 1981) (stating that the “principle purpose” of the parties, if ascertainable, is given great weight).

¹⁹² See *id.* § 201(2) (explaining which party’s interpretation prevails in the event that the parties have attached different meanings to the same contract term).

¹⁹³ See, e.g., HUGH COLLINS, REGULATING CONTRACTS 171, 232–34 (1999) (arguing for contract interpretation that serves goals of cooperation, flexibility, and distributive justice); Eyal Zamir, *The Inverted Hierarchy of Contract Interpretation and Supplementation*, 97 COLUM. L. REV. 1710, 1777–88 (1997) (arguing that contract interpretation serves social goals).

¹⁹⁴ See, e.g., George Cohen, *Interpretation and Implied Terms in Contract Law*, in 2 ENCYCLOPEDIA OF LAW AND ECONOMICS 125 (Gerrit De Geest ed., 2d ed. 2011); Juliet P. Kostritsky, *Interpretive Risk and Contract Interpretation: A Suggested Approach for Maximizing Value*, 2 ELON L. REV. 109, 135–37 (2011) (arguing that judicial intervention in a contract dispute should not be precluded if court enforcement would be welfare improving); Lipshaw, *supra* note 179, at 1019 (“All seem to agree that the normative goal of contract law generally is to enhance economic welfare by maximizing the joint economic surplus that arises from the transaction.”); Posner, *supra* note 180, at 1583 (arguing that the goal of contract interpretation is to minimize transaction costs).

contracts in a welfare-maximizing manner, let alone succeed at doing so. Nor do we believe that courts have the kind of information necessary to advance this goal. Would parties, in the name of “efficiency,” want to engage in complex litigation over how to maximize joint welfare? Moreover, many interpretation disputes involve purely distributive aspects of the transaction, which (by definition) cannot be resolved via the welfare maximizing principle.¹⁹⁵ Thus, it is unlikely that the survey interpretation method would resolve disputes in a systematically less efficient manner.

There are other normative goals that courts are urged to advance in the course of resolving interpretation disputes, such as favoring weak parties, punishing drafting sloppiness, and promoting external societal interests. No doubt, yielding to survey results would make it harder to advance these goals. If courts, for example, want to favor the meaning given by a protected group, they would have to rely on carefully selected samples to elicit this meaning. Even when a common meaning is verified, courts may want to advance a different meaning for good reasons. Under the survey method, such normative preferences could no longer pretend to be “interpretation” and other doctrinal levers would have to be candidly relied upon to advance them.

Still, in a great majority of cases courts do search for the plain and ordinary meaning. And even if this plain meaning is only the baseline for further inquiry and normative interpretation, the survey method can help establish it. Courts can rely on a survey to show that two meanings are equally plausible, opening the door for second-order criteria to determine which party prevails.

Ultimately, any objection to the survey interpretation method on the basis that it would spell the end of normative contract interpretation assumes that normative interpretation is actually practiced—a quixotic assumption with little empirical basis. Courts have neither the information nor the motivation to use contract interpretation disputes as a regular platform for social reform, hence their declared commitment to interpreting the contractual text “as the mass of mankind would view it.” The survey method is therefore at odds, not with the ongoing practice of contract interpretation, but primarily with *academic hopes* for a normative enterprise of interpretation.

¹⁹⁵ Omri Ben-Shahar, *A Bargaining Power Theory of Default Rules*, 109 COLUM. L. REV. 396, 396 (2009).

D. *Simplifying the Contracts*

The survey interpretation method, we said, could lower litigation costs, because it would rely less on the interpretive discretion of courts. For this to be accomplished, surveys would have to focus primarily on the plain meaning of the text, without presenting respondents with rich accounts of context. This sparseness would make it easier for drafters to predict what contractual language would be held to mean in the event of litigation.

Survey respondents would likely be presented with a snippet of the relevant text of the agreement and only a minimal amount of surrounding information.¹⁹⁶ Respondents are not compensated, sophisticated, or patient enough to attentively read detailed descriptions of the background facts and long excerpts of the boilerplate. Besides, context-rich survey questions would defeat an important goal of the method. They would require courts to determine how much extrinsic evidence to put in front of the respondents. But a primary goal of the survey interpretation method is precisely to avoid the endless adversarial fights over the scope of evidence.

When surveys rely on large samples of random respondents, we envision a presentation of the interpretive questions in a relatively bare format. Brevity was indeed one of the primary constraints we adhered to in designing the pilot surveys reported in Part III, which were all directed at lay respondents. In such surveys, the contractual texts and the competing interpretations would be presented to respondents stripped of additional facts regarding such matters as extrinsic communications between the parties, the norms that permeate in the trade, or the course of dealing and performance surrounding this contract. It is possible, of course, to run surveys with rich information about context, but the parties may disagree about the context—for example, about what was said in the precontractual communications—and thus choosing to present an interpretation survey that exposes the respondents to disputed context may run the risk of the court discounting its probative weight. Ultimately, the utility of the survey interpretation method as a litigation-simplifying device would be enhanced if courts were to endorse relatively context-free surveys.

This relatively context-free approach to interpretation would have a desirable effect on the simplicity of contracts. Firms would

¹⁹⁶ In instances where consumers are given no contractual text, and only asked about their expectations, context and normative preferences are doing essentially all the work in guiding respondents' non-random responses. In some instances, these sorts of expectations about the terms of a contract differ substantially from the reality. *See, e.g.*, Boardman, *supra* note 39, at 1082–84 (documenting common misconceptions held by consumers about the scope of their homeowners insurance).

want to draft simpler contracts with less cryptic jargon so as to increase their ability to predict how interpretive battles would come out. In the era of digital contracting it has become all the more costless for drafters to err on the side of verbosity; for example, we recently read United Airlines's 35,000-word passenger contract—a paragon of the digital consumer contract, yet found some striking ambiguities.¹⁹⁷ Even firms that try heroically to use lay language fall back on technical and lengthy legalese when dealing with weighty matters.¹⁹⁸ These incentives for excess and complexity would be constrained if the language has to be understood by lay readers who determine its interpretation. Drafters of agreements would have to consider how average respondents not trained in legal subtlety would read and understand the terms in a future survey. No longer able to rely on professional judges riding interpretive canons to the rescue (while reading between the lines), drafters' best strategy would be to write clearly. Various regulatory efforts prompt firms to do just that,¹⁹⁹ but without giving real incentive. The survey method could create the desired incentives for simplicity.

That said, there is one countervailing effect that justifies a caveat. For reasons we have already explained, we think it is more methodologically defensible to have survey respondents read the relevant excerpts of contracts rather than the entire contracts. If our approach becomes law we might see firms lengthen contracts. Rather than writing broad, abstract boilerplate that addresses a great many contingencies, they could decide that they want to draft narrower but clearer clauses—and lots of them. Contract complexity and contract length both undermine the ability of consumers to read and comprehend the terms of their agreements at the time of contract formation. We believe the clarity benefits of our approach outweigh the potential contract length downsides, but that would ultimately be an empirical question to be assessed after implementation.

¹⁹⁷ See Omri Ben-Shahar & Lior J. Strahilevitz, *David Dao v. United: What Does The Airline Contract Say?*, FORBES (Apr. 14, 2017), <https://www.forbes.com/sites/omribenshar/2017/04/14/david-dao-versus-united-what-does-the-airline-contract-say/#2ee0894c18ad>.

¹⁹⁸ See BEN-SHAHAR & SCHNEIDER, *supra* note 13, at 124–26 (explaining the tension between simplicity of contractual language and full disclosure).

¹⁹⁹ For example, the Truth in Lending Act requires credit card issuers to disclose terms “clearly and conspicuously,” though there has been significant confusion over what sorts of disclosures will satisfy that standard. Brandon Mohr, *Who Decides Whether Clarity Is Clear?: An Analysis of TILA’s Clarity of Disclosure Requirement in Actions by Consumers Against Credit Card Companies*, 32 PACE L. REV. 188, 188–89 (2012).

In any event, we don't want to overstate the value of simplification. Only very few consumers ever read the boilerplate²⁰⁰—ratios so microscopic that it is naïve to imagine that shorter boilerplate would become widely read. Yet for the occasional readers, the stakes of the transaction are perhaps idiosyncratically high, and making it easier for this small group to understand their contracts without forcing them to retain counsel is a benefit. The benefits of simplification may be even larger when consumers turn to information intermediaries when entering contracts. Homebuyers and sellers, for example, typically seek advice of real estate agents rather than lawyers. Consumers rely on services that rate or grade contractual aspects like privacy and warranties. And they often read prominent statements regarding withdrawal rights and termination penalties. In these situations, making the language more readily comprehensible could have benefits even for people who do not bother to read the fine print.

Another aspect of simplification is predictability. The survey method would make the interpretation of a contract more predictable *ex ante*, because it is based on a method that can be replicated—and pre-tested. Currently, parties who seek greater predictability have to deploy contract language that has already been interpreted by a court in the relevant jurisdiction, but such case law often does not exist. Even if it exists, different judges may respond to precedents differently, so in the event of litigation a lot could depend on the identity of the trial court judge whose rulings will prove dispositive. The survey method liberates the parties from such constraints. We argue in Part V below that by employing the kinds of experiments we described in Part III, contracting parties may be able to field test the meaning of terms before the agreement, or after the agreement but before engaging in the bargained-for conduct, all to identify ambiguities and resolve them. Basing the meaning of a contract on the responses of a nationally representative sample of consumers lets contract drafters identify litigation risk and plan accordingly. Thus, rather than await the dispute and its uncertain resolution, parties can predict the legal outcome by applying the survey method privately.

Other interpretive tools provide far less predictability—interpretive canons are contradictory, and policy judgments about efficient terms are made amidst great uncertainty and require subjective judgments by courts that may vary based on judge's policy preferences. By contrast, there is evidence suggesting that survey responses to inter-

²⁰⁰ See BEN-SHAHAR & SCHNEIDER, *supra* note 13, at 107 (2014) (identifying obstacles to the effectiveness of disclosures, such as illiteracy and the overload problem).

pretive questions will be stable over time.²⁰¹ This added predictability afforded by the survey interpretation method could increase drafting flexibility and promote innovation. One of the striking features of contractual boilerplates is their rigidity and stickiness. In a legal regime that relies on court interpretation, the parties prefer to cut and paste existing clauses that were already interpreted by courts and have acquired clear and predictable meaning.²⁰² Switching to new language, even if more suitable to the evolving circumstances, introduces a new interpretation risk that would only be resolved later, through further rounds of litigation. Insurance companies, for example, often prefer to stick with existing boilerplate that was interpreted by courts against them rather than redraft the policies, because that language is no longer ambiguous.²⁰³ Survey interpretation—by allowing firms to resolve the uncertainty prior to litigation—can foster greater drafting flexibility.²⁰⁴

E. *Uncovering Sophisticated Meaning*

Our “as the mass of mankind” rationale for the survey interpretation method envisions a consumer contract, drafted unilaterally by a sophisticated party and disseminated to many lay parties. But a variant of the method is potentially well suited to more sophisticated contracting environments, like merchant-to-merchant contracts and negotiated agreements between firms. Here, too, courts are often at a disadvantage relative to surveyed parties. Unlike the consumer context, where courts are more sophisticated than the transactors, in the

²⁰¹ See Matthew B. Kugler & Lior Jacob Strahilevitz, *The Myth of Fourth Amendment Circularity*, 84 U. CHI. L. REV. 1747, 1760 (2017) (explaining that the public’s beliefs about privacy remain relatively stable over time, despite major changes in the law).

²⁰² See generally GULATI & SCOTT, *supra* note 96, at 33–44; Omri Ben-Shahar & John A. E. Pottow, *On the Stickiness of Default Rules*, 33 FLA. ST. U. L. REV. 651 (2006) (explaining the appeal of using default rules rather than opt-out provisions).

²⁰³ See Michelle E. Boardman, *Contra Proferentem: The Allure of Ambiguous Boilerplate*, in *BOILERPLATE: THE FOUNDATION OF MARKET CONTRACTS* 176, 179 (Omri Ben-Shahar ed., 2007) (“Not only does past language become clearer over time in the insurer’s eyes, but the cost of each clause also becomes increasingly clear as actuarial data is collected and pooled. Changing language, even in an effort to decrease coverage, could be more costly.”); Boardman, *supra* note 39, at 1105–06 (explaining that drafting new language is risky for insurers because there is no guarantee that it will be read as intended).

²⁰⁴ Some contracts are aimed not only at the parties to them but to third parties as well. For example, the contract between an owner of vacant commercial property and a construction firm might affect various aspects of the relationship between the construction firm and its subcontractors on the project. Other contracts, like home mortgage agreements, may be written with an eye towards subsequent securitization. In these instances, we would expect that the sophisticated third parties (who are evaluating other contracts in an environment characterized by repeat play) will develop similar expertise to that developed by the sophisticated contracting parties themselves.

business-to-business context courts may be insufficiently fluent in the sector's lingo. They could benefit from the crowd-sourced knowledge that a survey would furnish.

In these settings, the universe of potential survey respondents has to be adjusted to include merchants, lawyers, and other professionals—respondents who have more insight as to the meaning of the contract. Suppose a dispute arises in the construction or advertising industries. Insiders will not be hard to find. Admittedly, as compared with consumer contracts, it would be harder to recruit neutral respondents (especially if many “insiders” have a stake in how the interpretive battle is resolved), and the size of the survey would have to be smaller. One might worry that respondents in merchant markets will provide self-serving answers in an attempt to steer legal standards to their favor. Hopefully, a clever design of the survey instrument could defeat this bias by obscuring the stakes and using random variation in phrasing to make it less obvious to respondents who would benefit from a particular interpretation. It would surely do better than the alternative methods of learning about trade usages and interpretations, which all too often rely on testimony by insiders who are frequently parties to the litigation.²⁰⁵

Interpreting negotiated contracts would require surveying lawyers—probably the only professionals working on such deals that understand the meaning of the legal terms. The survey would have to assemble a reasonably large number of transactional lawyers and gather their interpretations. This would not necessarily entail paying each lawyer their billable rate. For example, a firm seeking to resolve ambiguities might offer a free CLE program to lawyers in which their interpretations of text are initially collected. Then, the lawyers running the CLE would inform the lawyers in attendance of what consensus interpretations emerged from the survey and what, if anything, the relevant precedents say about similar language. Although one might like to suppose that lawyers follow the latest contract interpretation decisions closely, some evidence suggests that lawyers' suppositions when drafting agreements can be sticky and out of step with what courts are doing,²⁰⁶ suggesting that reliance on the survey interpretation method may have many of the benefits described above as well.

²⁰⁵ See Lisa Bernstein, *Custom in the Courts*, 110 Nw. U. L. REV. 63, 78–79 (2015) (explaining that among a set of cases studied, the most common type of usage evidence introduced was the testimony of a party or a party's employee).

²⁰⁶ Coyle, *supra* note 95, at 639 (explaining that drafting new language is risky for insurers because there is no guarantee that it will be read as intended).

Although we think of consumer contracts as our core case, there are significant advantages to interpretive consistency across contract types. Making surveys relevant for contracts of all sorts takes some pressure off the determinations that courts would have to make in boundary cases, such as employment agreements and insurance contracts that are regulated by the state. That said, because different contexts call for different kinds of surveys, some judgment calls would persist, and uncertainty about how those judgments would be resolved will impose costs on the parties. The relevant question, of course, is how these costs compare to those engendered by the substantial uncertainty that presently exists in contract interpretation.

V

THE DOCTRINAL FOUNDATIONS OF THE SURVEY INTERPRETATION METHOD

Part IV argued that the survey interpretation method presented in Part II and tested out in Part III has distinct desirable attributes. We argued that abandoning the present method of interpretation, which depends on jurists interpreting language rather than relying on the understanding of those the contract is intended to bind, and replacing it with surveys would simplify litigation and align contractual meaning with reasonable expectations.

Our exercise so far highlighted the difference between the proposed survey interpretation method and the traditional interpretation jurisprudence. In this Part, we reverse the lens. Rather than showing how different the two regimes are, we show some common DNA. We argue that the survey interpretation method is consistent with several principles and doctrines of contract law. We do this in two steps. First, we identify traces of openness in courts to the survey method, which suggest that the method's expansion to contract interpretation might be embraced. Second, we argue that contract law permits parties to opt into the survey method by agreeing in the contract that any future interpretation dispute would be resolved by surveys.

A. *The Survey Method in Courts*

To imagine how courts would view the survey method in contract interpretation, it is worth recalling how the method became so prominent in trademark litigation. As we noted at the outset, efforts to introduce surveys into evidence initially ran aground of the hearsay exception²⁰⁷ and judicial skepticism as to whether such surveys could

²⁰⁷ See *Elgin Nat'l Watch Co. v. Elgin Clock Co.*, 26 F.2d 376, 377–78 (D. Del. 1928) (rejecting lawyer's attempt to introduce an expert-conducted consumer survey on the basis

be accurate.²⁰⁸ Within a few decades, the opposition to consumer survey evidence had been overcome.²⁰⁹ And whatever concerns judges may have initially had about the use of online surveys in trademark litigation dissolved quickly enough²¹⁰—so much so that it is now considered an abuse of discretion to exclude even an online survey on the grounds that it does not “replicate real world conditions.”²¹¹ The use of surveys, and in particular low-cost online surveys, is so widely accepted (and expected) that courts no longer bother to justify this practice.²¹²

It is easy to see why surveys so naturally fit trademark and advertising law litigation. The fundamental question in these cases is how the intended lay audience understood and interpreted a particular communication. Since the goal is to figure out consumers’ perceptions and beliefs, what better way than to ask them?²¹³ It might be thought that interpreting a contract is unlike the inquiry into a likelihood of consumer confusion. Interpretation tries to track the meaning that the parties to *this* contract gave to the language they used, not necessarily the meaning that survey respondents would give. Moreover, interpretation is widely thought of as a legal filter that does more than decipher lay meaning; it is a tool for courts to regulate incentives to draft, comply with, enforce, and litigate contractual rights.²¹⁴ And indeed, so

of the hearsay rules, insisting that the court could only accept evidence from consumers who would testify personally in open court).

²⁰⁸ See *Du Pont Cellophane Co. v. Waxed Prods. Co.*, 6 F. Supp. 859, 884–85 (E.D.N.Y. 1934) (holding that the survey is not “competent, material, or relevant”), *modified on other grounds*, 85 F.2d 75, 80 (2d Cir. 1936).

²⁰⁹ See *Zippo Mfg. Co. v. Rogers Imps., Inc.*, 216 F. Supp. 670, 680, 686 (S.D.N.Y. 1963) (admitting a consumer survey as evidence in a trademark case).

²¹⁰ See, e.g., *Bach v. Forever Living Prods. U.S., Inc.*, 473 F. Supp. 2d 1110, 1115–16 (W.D. Wash. 2007) (holding that defendants may challenge the reliability of the survey, but not its admissibility).

²¹¹ *Fortune Dynamic, Inc. v. Victoria’s Secret Stores Brand Mgmt.*, 618 F.3d 1025, 1037–38 (9th Cir. 2010).

²¹² See *POM Wonderful LLC v. Organic Juice USA, Inc.*, 769 F. Supp. 2d 188, 199–201 (S.D.N.Y. 2011) (admitting surveys despite alleged methodological flaws). The leading trademark law treatise says flatly that, “[i]f a survey conducted using the internet is performed according to the generally accepted principles of surveying, then it should be admitted into evidence and evaluated the same as any other survey results.” MCCARTHY, *supra* note 43, at § 32:165.25.

²¹³ See Robert C. Bird, *Streamlining Consumer Survey Analysis: An Examination of the Concept of Universe in Consumer Surveys Offered in Intellectual Property Litigation*, 88 TRADEMARK REP. 269, 270 (1998) (“A consumer survey is a scientific method of presenting evidence of mental associations of a given group of people . . .”).

²¹⁴ See Omri Ben-Shahar, *The Tentative Case Against Flexibility in Commercial Law*, 66 U. CHI. L. REV. 781, 783–84 (1999) (analyzing how the U.C.C.’s non-formalist interpretive approach shapes the business practices of contracting parties); Eric A. Posner, *The Parol Evidence Rule, the Plain Meaning Rule, and the Principles of Contractual Interpretation*, 146 U. PA. L. REV. 533, 533 (1998) (stating that a court should apply a default or “gap-

goes the objection, surveys are rarely if ever used in contract litigation for precisely these reasons.

Not so fast. Courts increasingly think that surveys are useful even beyond the likelihood-of-confusion context. Under Rule 703 in the Federal Rules of Evidence, surveys are admissible whenever their presentation has a “probative value [in helping the jury evaluate the opinion that] substantially outweighs any prejudicial effect.”²¹⁵ It is this probative effect that is prompting courts to be increasingly receptive to, and even to demand, survey evidence. For example, a frequent question that arises in insurance contract litigation is the following: What reasonable expectations did a promise evoke?²¹⁶ Contracting parties claim that their reasonable expectations ought to be enforced—and courts willing to enforce such expectations are increasingly calling for surveys to establish the existence and content of the expectations.²¹⁷ Or, under the doctrine of unconscionability, courts have to determine, among other things, whether a term is surprising and unexpected to the typical consumer.²¹⁸ Under the American Law Institute’s proposed Restatement of Consumer Contracts, this can be shown by survey evidence.²¹⁹

Surveys are also useful in administering another contract interpretation doctrine—trade usage—that requires courts to interpret contracts and fill gaps by looking at practices regularly observed by transactors in the relevant market.²²⁰ It may not always be true that a trade usage exists. But what better way to prove its existence and content than by a survey of parties who ordinarily participate in the rele-

filling rule” that either maximizes the contract’s ex-ante value or provides ex-ante incentives to disclose sufficient information); Steven Shavell, *On the Writing and the Interpretation of Contracts*, 22 J.L. ECON. ORG. 289, 301 (2006) (demonstrating that it may be optimal not to enforce the “literal” (plain) meaning of a term, so as to save parties writing costs).

²¹⁵ See, e.g., *United States v. H & R Block, Inc.*, 831 F. Supp. 2d 27, 30 (D.D.C. 2011).

²¹⁶ See generally Robert H. Jerry II, *Insurance, Contract, and the Doctrine of Reasonable Expectations*, 5 CONN. INS. L.J. 21 (1998) (exploring the doctrine of reasonable expectations).

²¹⁷ See *Walker v. Life Ins. Co. of the Sw.*, No. CV 10-9198, 2015 WL 10761819, at *26 (C.D. Cal. Apr. 14, 2015) (holding that consumer surveys could have been, but were not, used in establishing consumer expectations arising from life insurance policies).

²¹⁸ See *O’Donovan v. CashCall, Inc.*, 278 F.R.D. 479, 500 (N.D. Cal. 2011) (stating that in evaluating unconscionability under California law, courts have to determine “the extent to which a provision is ‘hidden’ or unexpected”).

²¹⁹ See RESTATEMENT OF CONSUMER CONTRACTS § 5 cmt. 6 (AM. LAW INST., Council Draft No. 3, 2017).

²²⁰ See U.C.C. § 1-205 (AM. LAW INST. & UNIF. LAW COMM’N 1977) (stating that usage of trade, a practice “having such regularity of observance in a place, vocation or trade as to justify an expectation that it will be observed with respect to the transaction,” should give meaning to the terms of an agreement).

vant commerce? If parties contract over “Grade A” chicken,²²¹ a “breeding” bull,²²² or a “sound” horse,²²³ but later dispute what quality these adjectives entail, the survey method could harness the collective experience of merchants buying and selling chicken, bulls, or horses. If a clear majority interpretation emerges and if the sincerity of respondents can be trusted, the result should dictate—or at least inform—the court’s finding.

This is also the view of the U.C.C. “The existence and scope of such a usage are to be proved as facts,” says the U.C.C.,²²⁴ and courts have thus explicitly endorsed and encouraged the use of proof methods akin to surveys.²²⁵ But these have not always been large surveys. Instead, a common method of proving trade usage is experts’ testimony, usually qualitative accounts by members of the trade.²²⁶ This is a type of survey, but of a small *n*. Less common but more reliable is statistical evidence about the regularity of conduct—“frequencies of a given behavior in the trade”—elicited via formal surveys of merchants.²²⁷ To be sure, Lisa Bernstein finds that courts do not resort to rigorous survey evidence regarding the content or scope of trade usages. But she views the nonuse of surveys as an unfortunate failure of the courts to implement the most appropriate and probative methodology.²²⁸

²²¹ *Frigalment Importing Co. v. B.N.S. Int’l Sales Corp.*, 190 F. Supp. 116, 118 (S.D.N.Y. 1960).

²²² *Sherwood v. Walker*, 33 N.W. 919, 923 (Mich. 1887).

²²³ *Smith v. Hughes* (1871) 6 LRQB 597 at 606 (Eng.).

²²⁴ U.C.C. § 1-205(2).

²²⁵ *Compare* *Lion Oil Trading & Transp., Inc. v. Statoil Mktg. & Trading (US) Inc.*, No. 08 Civ. 11315, 2011 WL 855876, at *2 (S.D.N.Y. Feb. 28, 2011) (using a “survey of crude oil traders” to determine a trade usage), *and* *Timeline, Inc. v. ProClarity Corp.*, No. C05-1013, 2007 WL 1574069, at *8 (W.D. Wash. May 29, 2007) (endorsing, implicitly, the use of survey evidence in contract interpretation by criticizing an expert trade witness for failing to “identify any surveys or research he has conducted regarding the meaning or usage of these terms”), *with* *M.S. Distrib. Co. v. Web Records, Inc.*, No. 00 C 1436, 2003 WL 21087961, at *12 (N.D. Ill. May 13, 2003) (dismissing survey evidence because the survey’s methodology was amateurish).

²²⁶ FARNSWORTH, *supra* note 11, at 471 (noting that under the U.C.C., “[a] party commonly shows a usage by producing expert witnesses who are familiar with the activity or place in which the usage is observed”); GREGORY M. TRAVALIO ET AL., *NORDSTROM ON SALES & LEASES OF GOODS* 244 (2d ed. 2000) (“[P]resumably expert testimony will be necessary to establish a trade usage.”).

²²⁷ Jody S. Kraus & Steven D. Walt, *In Defense of the Incorporation Strategy*, in *THE JURISPRUDENTIAL FOUNDATIONS OF CORPORATE AND COMMERCIAL LAW* 193, 213 (Jody S. Kraus & Steven D. Walt eds., 2000).

²²⁸ *See* Bernstein, *supra* note 205, at 93 (arguing that without a “social scientific survey . . . it is unclear how the general subjective understanding of transactors across a relevant market could be reliably established.”).

Reliable evidence about the content of a custom or behavioral regularity requires that the people whose opinion is solicited by the court are representative of the sector of potential parties. In trademark and false advertising cases, experts are careful to survey consumers who say they are in the market for the goods or services at issue.²²⁹ The survey interpretation method for contracts would do likewise, surveying only those who are potential buyers of camping equipment in a dispute over the meaning of a North Face sleeping bag warranty, and surveying a sample of everyone with Internet access in a dispute over the meaning of Gmail's terms of service. People in a position to buy homes, not realtors or mortgage lenders, should interpret residential mortgage notes, and drivers should interpret auto insurance policies. If this basic alignment is preserved, the survey method would be entirely consistent with the widely-accepted goal of interpretation—to identify the meaning that *these* parties probably gave to the contract.

* * *

The survey method is a novelty, but it seeks to validate the basic principles that underlie traditional interpretation practices. Introducing survey evidence in contract litigation therefore does not require a retooling of contract doctrine. It requires gradual openness to more forms of probative evidence. We speculate that adoption of the survey method would thus proceed incrementally. Initially, courts might be willing to examine survey evidence as persuasive, not conclusive, evidence. They might rely on survey results to further strengthen the position towards which they already lean, and perhaps help justify summary judgments. Why ignore such input altogether? We also think the survey method is most likely to be adopted in major class action suits like the *Google* litigation.²³⁰ The stakes would be high enough to make it sensible for a litigant to invest in an innovative argument, one that could potentially prove decisive in a complicated case. After such early successful adoptions, the method could spread, just as it did in trademark and false advertising law.

²²⁹ This is known as the “relevant universe of consumers.” See generally Shashank Upadhye, *Trademark Surveys: Identifying the Relevant Universe of Confused Consumers*, 8 FORDHAM INTELL. PROP. MEDIA & ENT. L.J. 549 (1998) (explaining how proper identification of the relevant universe of consumers is critical to the success of a trademark holder's infringement case).

²³⁰ *In re Google Inc. Gmail Litig.*, No. 13-MD-02430, 2013 WL 5423918 (N.D. Cal. Sept. 26, 2013).

B. *Opting into the Survey Interpretation Method*

The previous section explored several threads that may connect the survey interpretation method with existing interpretive doctrines. Other than the trade usage and reasonable expectations jurisprudence, these connections are admittedly tentative. While we would like courts to appreciate the value of surveys in providing reliable interpretation of contracts, so far they have not done so systematically, in part because lawyers have so far not been encouraging them to consider survey-related innovations.

The question we now ask is whether courts may be instructed to use survey interpretation by the parties. Could parties opt into the survey method? We explore this possibility in two steps. First, we argue that interpretation doctrine is a default rule that can be contracted around. Second, we identify the rich opportunities that such customization of interpretation rules might accord the parties.

1. *Customizing Interpretation Law*

Are interpretation doctrines default rules? In one sense, the answer must be yes: Interpretation comes into play only when the contract meaning is not explicit. Drafting clear and explicit meaning would therefore always override interpreted meaning. But such drafting would be an overwhelming task, given the endless potential gaps in contractual language.²³¹ Can the parties, instead, override contract law's interpretation canon by changing the rules? Can they write in their contract specific instructions for courts to resolve ambiguities using different techniques than the ones courts would otherwise deploy, and specifically to apply the survey method?

Such opting out of interpretation rules could surely occur *ex post*, at the time of the dispute resolution. The parties could agree and instruct the court to rule out some interpretations or consider only some interpretive clues. If courts are resistant, the private override could be achieved through a stipulation of fact or even settlement. Accordingly, the parties could agree to conduct a survey and live by its results. Indeed, parties use surveys outside the courtroom to settle trademark cases. A study of practitioners reports that "surveys are used heavily in pretrial assessments and strategic decision making," and play "key roles in claim evaluation and are understood by attorneys as an influential settlement tool for both sides."²³²

²³¹ See Robert E. Scott & George G. Triantis, *Anticipating Litigation in Contract Design*, 115 *YALE L.J.* 814, 838–48 (2006) (analyzing the choice of vague or precise contract terms).

²³² Diamond & Franklyn, *supra* note 44, at 2061.

More interestingly, opting out of the default interpretation rules could also be attempted *ex ante*, at the time of entering the contract. The parties could include in their contract an “interpretation clause” instructing the court to interpret the text via a specific survey mechanism. Under our approach, such a clause would need to be drafted in a sufficiently clear way to make its meaning comprehensible to most laypeople who took the time to read the language at issue. While hardly a piece of cake, we think this goal is achievable.

Courts generally follow *ex ante* instructions on process, especially when they are not perceived to conflict with mandatory protections.²³³ For example, parties are allowed to override the default rules of interpretation by explicitly drafting “entire agreement” or “merger” clauses instructing courts to ignore other parol agreements (that might otherwise be looked at), and courts ordinarily comply.²³⁴ Or, parties may draft “anti-waiver” or “no oral modification” clauses instructing courts to ignore subsequent patterns of performance in interpreting the meaning of a term, and again courts largely comply.²³⁵ Subject to standard protective constraints—mostly addressing asymmetry between the parties and anti-deception goals—courts are cautiously willing to set aside their interpretive strategies and follow the ones the parties chose *ex ante*.²³⁶

Would this permissive approach extend to contractual clauses instructing courts to utilize survey evidence in interpreting the contract? In principle, we see no reason to exclude such agreements. The entire purpose of contract interpretation is to fulfill the parties’ express intent. It would be paradoxical if, in the course of carrying out this fulfillment mission, courts would blatantly ignore the parties’

²³³ See generally Alan Schwartz & Robert E. Scott, *Contract Theory and the Limits of Contract Law*, 113 *YALE L.J.* 541, 548, 609–18 (2003) (suggesting that courts treat interpretation rules as mandatory but advocating a permissive regime instead).

²³⁴ See 1 JAMES J. WHITE & ROBERT S. SUMMERS, *UNIFORM COMMERCIAL CODE: PRACTITIONER TREATISE SERIES* 182 (5th ed. 2006) (stating that “merger” clauses are generally valid and effectively preclude the introduction of extrinsic evidence at trial); see also FARNSWORTH, *supra* note 11, at 423–24 (explaining that parties can draft “merger” or “entire agreement” clauses to alleviate uncertainty and ensure that the contract will be treated as an integrated agreement).

²³⁵ See U.C.C. § 2-202 cmt. 2 (AM. LAW INST. & UNIF. LAW COMM’N 1977) (stating that course of dealing and trade usage are to be used as interpretive tools, “[u]nless carefully negated” in the contract). *But see* Omri Ben-Shahar, *supra* note 214, at 790–91 (noting that courts sometimes read anti-waiver provisions as themselves subject to modification by course of performance).

²³⁶ See RESTATEMENT OF CONSUMER CONTRACTS § 8(c) (AM. LAW INST., Council Draft No. 3, 2017) (The requirement that courts follow interpretation guidelines that appear in the standard contract terms may be set aside “when the standard contract terms contradict or unreasonably limit an affirmation or promise, which is made part of the basis of the bargain between the business and the consumer.”).

desire to be governed by the survey interpretation method. We also see no reason why the courts ought to respect parties' instructions with respect to the substantive bargain and not with respect to these process issues. In fact, courts generally enforce parties' procedural instructions. They enforce choice of forum and arbitration clauses, limited discovery, statute of limitations clauses, enhanced requirements of writing or other proof, and more.²³⁷

Contracting into the survey method of interpretation is an opportunity for the parties to specify not only the general preference for the survey method, but also the precise survey procedures. Such *ex ante* specification would help overcome a potential problem with the method that sometimes plagues trademark litigation (where *ex ante* agreement over litigation procedures is impossible)—the battle of the surveys. For example, the parties may specify which survey firm would conduct a survey and how the cost would be allocated.²³⁸ Likewise, the parties may contract for an arbitration forum that is committed, as part of its guaranteed procedures, to deploying the survey interpretation method.

2. *Pretesting Contract Language*

If parties can anticipate the use of the survey method in adjudication, they may also use it to pretest the language of their contract. As

²³⁷ A substantial literature examines contracting over procedure. *See, e.g.*, Kevin E. Davis & Helen Hershkoff, *Contracting for Procedure*, 53 WM. & MARY L. REV. 507 (2011); Jaime Dodge, *The Limits of Procedural Private Ordering*, 97 VA. L. REV. 723 (2011); Daphna Kapeliuk & Alon Klement, *Changing the Litigation Game: An Ex Ante Perspective on Contractualized Procedures*, 91 TEX. L. REV. 1475 (2013); Michael L. Moffitt, *Customized Litigation: The Case for Making Civil Procedure Negotiable*, 75 GEO. WASH. L. REV. 461 (2007); Henry S. Noyes, *If You (Re)Build It, They Will Come: Contracts to Remake the Rules of Litigation in Arbitration's Image*, 30 HARV. J.L. & PUB. POL'Y 579 (2007); Judith Resnik, *Procedure as Contract*, 80 NOTRE DAME L. REV. 593 (2005); Robert J. Rhee, *Toward Procedural Optionality: Private Ordering of Public Adjudication*, 84 N.Y.U. L. REV. 514 (2009). For more skeptical takes, see David A. Hoffman, *Whither Bespoke Procedure?*, 2014 U. ILL. L. REV. 389 (2014), and W. Mark C. Weidemaier, *Customized Procedure in Theory and Reality*, 72 WASH. & LEE L. REV. 1865 (2015).

²³⁸ In opting into a survey interpretation regime, firms might be tempted to assign the interpretation to survey companies that favor these firms' interests, and survey firms might reach firm-friendly outcomes in the hope of winning future survey business. Since consumers do not scrutinize such terms in the contract, this risk is real and could undermine the integrity of the survey interpretation method. A similar phenomenon became widespread in consumer credit contracts, when firms included arbitration clauses that assigned dispute resolution jurisdictions to biased arbitrators. *See* Dan Slater, *San Francisco Sues Provider of Arbitrators, Alleging Bias*, WALL ST. J.: L. BLOG (Apr. 7, 2008, 9:23 AM), <http://blogs.wsj.com/law/2008/04/07/san-francisco-sues-provider-of-arbitrators-alleging-bias/>. As it was dealt with in the arbitration context, the risk of corrupt interpretation clauses would have to be addressed through doctrines like fraud and unconscionability.

Section IV.B. indicated, rather than wait for a dispute and take the risk of an adverse outcome in litigation, a cautious drafter of the language can verify its meaning by putting it to the same survey test that courts would employ *ex post*. Michelle Boardman even suggested, wisely, that a pre-tested proven meaning would factor into subsequent court battles.²³⁹ This admission would spare the parties the hassle of rerunning the survey.

There is, to be sure, an advantage in waiting for the dispute to arise. Advance testing is overinclusive, applying to clauses that may never get litigated or to contracts and transactions that may never be subject to litigation. This added cost of overinclusive pretesting has to be balanced against the benefit of reduced litigation risk. It is likely that novel clauses, or ones that regulate large stakes, would be pretested. The contract, recall, is part of the product. Many features of products are pretested prior to their release to markets; why not also the legal features?

The survey method should work well in helping contract drafters identify “known unknowns” when it comes to contractual meaning. Yet even with extensive pretesting, there will be some “unknown unknowns” for firms—provisions whose ambiguity the firm drafters could not anticipate, perhaps because of a subsequent and unforeseen technological development. With “unknown unknowns,” it will be impossible to design survey vignettes that appropriately test for the possibility of a subsequent ambiguity. There is not really any method that will reduce the uncertainty surrounding the interpretive language of “unknown unknowns,” a shortcoming that the survey method shares with every available alternative.

The incentive in favor of pretesting may be further bolstered if firms utilize standard boilerplate. A trade group, for example, may pretest an entire standard form contract, which would subsequently be utilized by numerous parties over a long time. The balance in favor of pretested meaning may also shift if a contract is part of a mass distribution of products or services. Google, for example, may wish to pretest many clauses in its user agreement. It could also publish the externally certified results of its pretesting if it wanted to signal to potential litigation opponents that it can introduce such data in the event of litigation. This would no doubt help fend off litigation.

The combination of private opt-in and pretesting could create a market for contract interpretation. In this market, survey firms would compete for the business of *ex post* interpretation of contract language and *ex ante* pretesting. These firms could become drafters of

²³⁹ Boardman, *supra* note 39, at 1099–101.

boilerplate, offering parties contract forms with known meaning (perhaps kept a trade secret), and could even warrant the meaning and insure against adverse legal interpretation. Competition could go a long way toward ensuring that surveys are done well. In this contracting ecosystem, the role of courts would be to guarantee that survey firms are not colluding with their clients and that interpretation clauses in contracts are not used to gut mandatory protections.

VI

FURTHER EXTENSIONS AND CONCLUDING THOUGHTS

Interpretation canons have been with us for centuries. The age-old Williston-Corbin debate over whether interpretation of contracts should be fact-intensive or narrowly textual has not been resolved,²⁴⁰ and there is no reason to think it ever will be. The law of interpretation will remain non-uniform, inconsistent, and costly if it continues along this path. Judges will bemoan, if not ridicule, the procedures that the common law of contracts imposes on them when they set forth to interpret contracts.²⁴¹

It is time to modernize the law of contract interpretation, and this Article attempts to make some first strides in that direction. Surveys are revolutionizing many social and economic institutions—from product ratings to Facebook “likes” and even classroom pedagogy. Why not litigation too? The rationale for using surveys in litigation is already well accepted in some areas, but the scope of the practice was previously limited because of the costs of conducting them. The advent of cheap, online, but still representative, survey technology is an opportunity to spread the method. Under our approach, the results of surveys and experiments would be conclusive with regard to interpretation disputes in some contexts (where a clear consensus exists) and relevant in others.

Although our focus here is on contract interpretation, the survey and experimental methods we discuss may be helpful in various other domains of litigation. In these last pages, we offer some preliminary thoughts on their potential extension to other domains, but also a criterion that we think ought to limit the scope of the method.

²⁴⁰ See Adam B. Badawi, *Interpretive Preferences and the Limits of the New Formalism*, 6 BERKELEY BUS. L.J. 1, 3–4 (2009) (discussing the debate between contextualists and formalists).

²⁴¹ We are reminded of Judge Kozinski’s complaint that California’s interpretation doctrine “casts a long shadow of uncertainty over all transactions negotiated and executed under the law of California.” *Trident Ctr. v. Conn. Gen. Life Ins.*, 847 F.2d 564, 569 (9th Cir. 1988).

One possible extension is to class actions. A major issue in class action litigation is the extent to which common issues of law or fact predominate over issues that affect class members as individuals and warrant treating a would-be class of plaintiffs as a collectivity with unified interests.²⁴² In consumer class actions, the claim often turns on a reading of the contract, and we already showed that its interpretation could be done via consumer surveys. But surveys could also resolve issues of commonality. For example, in some instances the language of the contract at issue may have changed over time, and courts have to figure out whether these modifications preclude a set of plaintiffs from sharing common issues. Courts considering these kinds of cases have tended to view variations in contracts as material and have denied class certification as a result.²⁴³

The survey-based approach offers a markedly more satisfying way to tackle the commonality problem. Just because consumers see different contract language, it does not follow that consumers' understanding of the underlying bargains would differ in any significant way. The Google and Yahoo experiments illustrate this emphatically.²⁴⁴ A survey showing that respondents share common reactions to different versions of the contractual language should override courts' assessments of the materiality of the revisions.

Moving beyond the realm of contract law, surveys could be used in a variety of other contexts where a legal determination is based on reasonable expectations—namely, on the question of how ordinary people view a particular issue. For example, under Fourth Amendment law, the question of what ordinary people expect with respect to police surveillance can be relevant or even dispositive, suggesting a valuable role for survey data of this kind.²⁴⁵ Or, in designing various disclosure laws, the question of how people understand the

²⁴² See *Tyson Foods, Inc. v. Bouaphakeo*, 136 S. Ct. 1036, 1045 (2016) (discussing the requirement under Federal Rule of Civil Procedure 23(b)(3) that common questions of law or fact predominate over individual factors).

²⁴³ See, e.g., *Sacred Heart Health Sys., Inc. v. Humana Military Healthcare Servs., Inc.*, 601 F.3d 1159, 1171–76 (11th Cir. 2010) (holding that significant variation in material terms overwhelmed common question at issue); *Broussard v. Meineke Disc. Muffler Shops, Inc.*, 155 F.3d 331, 340 (4th Cir. 1998) (holding that the commonality requirement for class action certification was not satisfied).

²⁴⁴ See *supra* Section III.B.5.

²⁴⁵ See Brief of Amici Curiae Empirical Fourth Amendment Scholars in Support of Petitioners, *Carpenter v. United States*, No. 16-402 (U.S. Aug. 14, 2017), 2017 WL 3530963 (making a similar argument in a pending Fourth Amendment case on behalf of Strahilevitz and fourteen other empirical scholars and drawing on the survey-based, scholarly literature that examines what level of geolocation privacy ordinary Americans actually expect); Matthew B. Kugler & Lior Jacob Strahilevitz, *Actual Expectations of Privacy, Fourth Amendment Doctrine, and Mosaic Theory*, 2015 SUP. CT. REV. 205 (2016) (arguing that survey data about what the public expects is relevant to Fourth Amendment doctrine).

disclosure is sometimes (although not always) critical to evaluating the adequacy of disclosure.²⁴⁶ Using surveys to resolve this question would be a good and important step forward, correcting unrealistic factual assumptions about the effects of disclosure, and perhaps ridding us of useless disclosure mandates.

Another potential extension of the survey methodology is to the adequacy of jury instructions. Appellate courts regularly evaluate how jurors would respond to variations in jury instructions,²⁴⁷ and an experimental literature has emerged to study these questions,²⁴⁸ but the judicial decisions tend to have a blind spot as to how their analysis might be improved with recourse to empirical evidence. For example, in some cases where there are variations between the underlying criminal statute and the instructions given to the jury concerning the application of the law, courts seize on these variations to reverse convictions.²⁴⁹ Or they hold such variations to be harmless error.²⁵⁰ They make these decisions without the benefit of survey evidence that could cast light on the question of whether the alterations or omissions from the statutory language—minor or major—change the way that lay jurors understand their charge.²⁵¹

Finally, it is tempting to think that if the survey method could help interpret contracts, it could be expanded further, to resolve other legal issues, such as statutory interpretation,²⁵² applications of legal

²⁴⁶ See, e.g., KLEIMANN COMM’N GRP., KNOW BEFORE YOU OWE: EVOLUTION OF THE INTEGRATED TILA-RESPA DISCLOSURES 29–34 (2012) (describing the methodology used to study how participants interacted with and interpreted the disclosures).

²⁴⁷ See, e.g., *United States v. Southwell*, 432 F.3d 1050, 1052–56 (9th Cir. 2005) (overturning verdict because the district court failed to answer jury question as to whether they could convict defendant if they did not agree unanimously on defendant’s sanity).

²⁴⁸ See, e.g., Dan Simon, *More Problems with Criminal Trials: The Limited Effectiveness of Legal Mechanisms*, 75 LAW & CONTEMP. PROBS., no. 2, 2012, at 167, 173–85 (analyzing the effectiveness of various types of instructions); Peter Miejes Tiersma, *Dictionaries and Death: Do Capital Jurors Understand Mitigation?*, 1995 UTAH L. REV. 1, 10–23 (discussing the results of empirical studies measuring jurors’ comprehension of jury instructions).

²⁴⁹ See, e.g., *United States v. Fernandez*, 722 F.3d 1, 16–27 (1st Cir. 2013) (holding that the jury instructions, which included language that did not track the statute, were erroneous).

²⁵⁰ See, e.g., *Tillman v. Cook*, 215 F.3d 1116, 1123–27 (10th Cir. 2000) (holding that the jury instruction correctly communicated the concept of reasonable doubt to the jury).

²⁵¹ See, e.g., *State v. Paul*, 104 A.3d 1058, 1059–62 (N.H. 2014) (insisting that jurors would appreciate the significance behind the choice of “must” in the first clause and “should” in the second clause—notably, that juries have the power to nullify criminal convictions—when told “[i]f you have a reasonable doubt . . . you *must* find the defendant not guilty,” but “if you find that the State has proved all of the elements of the offense charged beyond a reasonable doubt, you *should* find the defendant guilty”).

²⁵² See William Ortman, *Chevron for Juries*, 36 CARDOZO L. REV. 1287 (2015) (proposing a larger role for lay juries in statutory interpretation matters); see also J.P. Sevilla, *Measuring Ordinary Meaning Using Surveys* (Sept. 10, 2014) (unpublished

standards (for example, what constitute reasonable precautions), and even constitutional interpretation. This is not the conclusion our analysis advocates, and the arguments for such extensions have to be made separately. Our case for the survey method is based on the premise that respondents are the right people to ask about the meaning of a text that governs their behavior. We poll them because the primary question the court is trying to answer is: How do the recipients of the contractual texts and other communications understand them? Respondents are not usually the right people to ask what the Constitution says, or how to implement a negligence standard, because our legal system does not think that the proper answer to these questions necessarily turns on the ordinary lay understanding of a text.

working paper), <https://ssrn.com/abstract=2466667> (arguing that ordinary meaning of statutory language can be objectively measured using surveys).