Interpreting Contracts via Surveys and Experiments

Omri Ben-Shahar
Lior Strahilevitz
INTERPRETING CONTRACTS VIA SURVEYS AND EXPERIMENTS

Omri Ben-Shahar* and Lior Jacob Strahilevitz**

Abstract
Interpreting the language of contracts is 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. The article demonstrates the rich potential under this method to examine variations of the contractual language that could have made an intended meaning clearer. A similar survey regime has been applied successfully in trademark and unfair competition law to interpret precontractual messages, and the article shows how it could be extended to interpret contractual texts. To demonstrate the technique, the article applies the survey interpretation method to five real cases in which courts struggled to interpret contracts. It then provides normative, pragmatic, and doctrinal supports for the proposed regime.

* Leo and Eileen Herzel Professor of Law, University of Chicago.

** Sidley Austin Professor of Law, University of Chicago. The authors thank Ronen Avraham, Ruoying Chen, Gillian Hadfield, Jennifer Nou, Richard McAdams, Ariel Porat, Geof Stone, Eyal Zamir, and workshop participants at Keio University Law School, the University of Chicago, and Sidley Austin LLP for helpful comments, Taylor Coles for outstanding research assistance, and the Coase-Sandor Institute for Law and Economics for financial support. A special thanks is owed to Matthew Kugler for terrifically helpful assistance with data collection and analysis.
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’s 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. And from California. Witnesses would be marched in from all over. Everyone understood that the plaintiff (and defendant) would be 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 marks were apt 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 their having been deceived should not be dispositive.⁵ By the early 1960’s judicial opposition to consumer survey evidence had crumbled completely.⁶ Such evidence is now de rigueur in trademark and false advertising litigation.⁷

¹ Elgin Nat. Watch Co. v. Elgin Clock Co., 26 F.2d 376, 377-78 (D. Del. 1928).
² Du Pont Cellophane Co. v. Waxed Products Co., 6 F. Supp. 859, 884-85 (E.D.N.Y. 1934), modified but not on this ground, 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); see also Schering Corp. v. Pfizer Inc., 189 F.3d 218, 224-25 (2d Cir. 1999) (summarizing the development of the case law).
⁴ Beverly W. Pattishall, Reaction Test Evidence in Trade Identity Cases, 49 TRADEMARK REP. 145, 151 (1959).
⁵ Id. at 148-49.
⁷ Schering Corp., 189 F.3d at 225.
When it comes to contract interpretation, American law may as well be stuck in 1928. The question of what a consumer contract means and what expectations it arouses among consumers is not determined based on how hundreds or thousands of representative consumers surveyed 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 paper we offer a better alternative.

Interpreting the language of contracts is 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 though large 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 courts should admit to inform their interpretation of contractual language? New York and California present two polar extremes, with one admitting perhaps too little and the other surely too much. And within state jurisdictions, courts are insubordinate. The Illinois Supreme Court has instructed lower courts to consider only some information, but to no avail; lower courts do otherwise.

Second, contract interpretation law is in trouble because it is overly complex. Courts employ a mishmash of conflicting methodologies—including textual, formalist, purposive, and functional—to elicit the meaning of texts. They rely on various policy goals—accuracy, reduction of transaction and litigation costs, improve drafting, and reward information exchange—to expand or shrink the scope of their contextual

---

8 See, e.g., Mitchill v. Lath, 257 N.Y. 377 (N.Y. 1928) (Lehman, J., dissenting) (arguing that the application of New York’s parol evidence rule is too restrictive); Trident v. Connecticut General Life Ins., 847 F.2d, 564 (9th Cir. 1988) (arguing that California’s erosion of the parol evidence rule is too permissive).

inquiry. And they fall back on archaic canons of interpretation that have little to do with the way people read prose.

Third, and largely due to its inconsistency and complexity, contact 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. It is thus not surprising that, reading a lengthy boilerplate, only 9% of consumers understood that their contract had a legally enforceable mandatory arbitration clause.

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 millions of dollars fighting over what evidence to admit and what procedures to follow in interpreting the contract. 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 override more boldly entire codes of default rules. Contracts are assuming a more prominent role in regulating market transactions, and their interpretation is becoming a battle of greater stakes. The twentieth century question “what is ‘chicken?’” has become the twenty-first century question of whether Google is contractually permitted to scan its users’ emails.


13 See, e.g., Shareholder Representative Serv. LLC v. Airbus Americas, Inc., 791 S.E.2d 724, 727 (Va. 2016) (a case involving the breach of a merger agreement in which the court awarded the prevailing party attorney’s fees of nearly $3.9 million); Truserv Corp v. Morgan’s Tool & Supply Co., Inc., 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).


This paradoxical combination, of 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. We propose such a process—a survey interpretation method. Instead of asking judges and juries to interpret contracts, the meaning of disputed contractual clauses would be determined by polling a large representative sample of disinterested respondents. Let the majority of survey respondents decide.

The survey interpretation method might appear 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 lodge it on several supporting pillars. As our prologue revealed, a very similar method is widely used with satisfactory results in resolving interpretation disputes over closely related texts—appearing in advertising messages and descriptions of products. The disputing parties in such cases (primarily 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 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 some of the key methodological issues. We bolster the pragmatic argument by running a pilot study in which we apply the survey method to resolve actual interpretation disputes. Specifically, we picked several real cases in which courts struggled to interpret contracts coherently and we conducted large-scale surveys asking respondents to interpret the same language. We were able to compare the results of the traditional and the survey methodologies, and we highlight in the article the attractive features of the latter. We also used the survey method to test whether refinements in the contract language shown respondents elicit different majority interpretations. We did this to identify potential biases of respondents, but also to compare the contractual text with variations that might improve its clarity. Such results—and especially the proven existence of (unused) clearer language—allow us to provide a more rigorous and first-of-its-kind foundation for “whose meaning prevails” doctrines like interpretation-against-drafters and mutual mistake.

The second defense is normative. The survey method advances a particular conception of meaning: attaching to contracts the understanding those to whom they are written assign. (We discuss in the article why this, and not other criteria of meaning, is appropriate.) 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, investors) are interpreted by surveys of sector members. Beyond loyalty to meaning, the survey interpretation method has additional desirable qualities. One quality is the incentive to draft short, simple, and widely understandable contractual texts ex ante, to reduce the risk of misinterpretation. Perhaps such “thin” contracts would continue to go unread by most people, but aspiring readers would at least stand a chance. Another quality is low litigation cost. In the era of online panels, the cost of resolving interpretations 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 contracts are predictable, because they can be pre-tested. 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 central 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 in advance the survey company. This would also enable the parties to put the language of the contract to the test, ex ante. Companies 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, the article is structured as follows. Part I states the problem: existing contract interpretation law is inconsistent and complex, and it generates too much litigation at too high a cost. Part II introduces the solution—the survey interpretation method, showing 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

... Interpretation 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. Atiyah\(^\text{16}\)

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 would 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 would a typical person entering such a contract understand it—what expectations this language evokes among ordinary parties.\(^\text{17}\)

Words, though, rarely have singular meaning. No matter how carefully drafted, the reasonable understanding of a text 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 also stewing chicken?\(^\text{18}\) 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 on route from Miami to Puerto Rico, in international waters?\(^\text{19}\)

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, past and concurrent dealings, and trade norms may be presented.


\(^{17}\) Farnsworth, supra note 10, at 440.


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 dictionary meaning of contract text, and examine in addition the context in which the language was used, to determine whether more than one meaning in 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\(^\text{20}\)—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 (which we will call the problem of inaccuracy). 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.\(^\text{21}\) This fact-intensive approach 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. Similar to the plain meaning rule, the parol evidence rule establishes whether the written contract integrates the entire agreement or can be supplemented by reference to prior understandings. Here, too, the judge has to decide whether to look only internally at the four corners of the contract, or to admit additional extrinsic evidence.\(^\text{22}\) Likewise, the doctrine of reasonable expectations allows the court to bridge between the perfunctory letter of the contract and its spirit by enforcing a meaning consistent with the intentions of the parties as evidenced by other surrounding circumstances.\(^\text{23}\) In most states, courts apply such expansive interpretation only after determining that the text itself is ambiguous and susceptible to the additional meaning.\(^\text{24}\) These courts face the same methodological challenge: is a term ambiguous? How much extrinsic evidence should be employed in determining whether it is?


\(^{22}\) Farnsworth, supra note 10, at 422.

\(^{23}\) Restatement (Second) of Contracts § 211(3); C&J Fertilizer Inc. v. Allied Mutual Ins. Co., 227 N.W.2d 169 (Ia. 1975).

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 satisfactory. 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.\(^{25}\) Interpretation is a task that requires judges to apply intuition and common sense, namely mental processes that vary with the judges’ backgrounds and experiences.\(^{26}\)

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 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,”\(^{27}\) without contextual baseline. (Some information is invoked, for example, to know that “Grade A chicken” refers to supply chain standards and not to hygiene score of a Chick-Fill-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 pretrial stage, would be required to examine all extrinsic evidence.

---


26 See A L Corbin, *3 Corbin on Contracts* §535, 17–18 (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.”)

27 Orman v. Stanray Corp. 437 F.2d 231, 234 (7th Cir. 1971).
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 datum 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 textual approach of the four corners text, lower courts often follow the contextual approach, with occasional approval by the Seventh Circuit. One cannot shake the cynical impression that the choice of approach—more versus less information—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). 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.’”

A fundamental unresolved distinction 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 fits poorly 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—

---

28 See, e.g., Air Safety, Inc. v. Teachers Realty Corp. 706 N.E.2d 882 (Ill. 1999); Gallagher v. Lenart, 874 N.E.2d 43 (Ill. 2007).
29 See CONTRACT LAW, supra note 9, at §§ 6-16, 6-17.
30 FARNSWORTH, supra note 10, at 464.
a judge should defer to a jury only when the meaning depends on the credibility of, or the choice among reasonable inferences from, extrinsic evidence. But courts are far from uniform in their treatment of this distinction.

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 forcing 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 American jurisdiction has a sound basis for characterizing its contract law in that way.

II. The Survey Interpretation Method

This section presents a novel approach to interpretation of contracts. To determine the meaning of a text, courts ought to rely on surveys. A survey would ask a pool of respondents that resemble the contracting parties what meaning they assign to the language. If the respondents are split evenly, the term should be regarded as ambiguous. But if one meaning garners greater support among the survey respondents, it presumptively ought to prevail.

There are flickers of reception of this idea in contract doctrine, and we discuss them in Section V. A couple of law review articles have flagged the idea of a more empirically inclined approach to contract interpretation. And (then Judge) Sotomayor

32 Restatement (Second) of Contracts § 212(2) (1981);
33 Farnsworth, supra note 10, at 477-8.
34 Most notably, see 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, the salience of which would then be increased for consumers); and Ariel Porat & Lior Jacob Strailevitz, Personalizing Default Rules and Disclosure with Big Data, 112 Mich. L. Rev. 1417 (2014) (advocating the use of surveys to identify the contractual default terms that would be offered to demographically similar consumers). Another paper empirically measured consumers’ understandings of the likelihood of their being able to redeem frequent flier miles to see whether those understandings were accurate. That paper suggests that there is a contrast between expected redemption rates and realized redemption levels, and posits that this variance might suggest that the agreements are contracts of adhesion, but the study doesn’t focus on consumers’ interpretation of contract or policy text. See Ann Morales Olazabal et al., Frequent Flyer Programs: Empirically Assessing
has noted that while the 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 Equal Educational Opportunities Act.35 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 been relying on surveys.36 In a typical trademark infringement 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. 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.37 Their use in competitors’ disputes to decipher how people interpret communications is de facto “black letter law.”38

The gist of the approach presented in this section is to extend the existing survey methodology, currently utilized to interpret important precontractual communications,

References:

35 Schering Corp. v. Pfizer Inc., 189 F.3d 218, 225 (2d Cir. 1999) (Sotomayor, J.) (citing Keith v. Volpe, 858 F.2d 467 (9th Cir. 1988) and Debra P. v. Turlington, 730 F.2d 1405 (11th Cir. 1984)).


37 6 J. THOMAS McCARTHY, IN McCARTHY ON TRADEMARKS AND UNFAIR COMPETITION § 32:195 (4th ed. 2016) (plaintiff’s failure to present a consumer survey creates the perception that it is “less than deadly serious about its case.”); Sandra Edelman, Failure to Conduct A Survey in Trademark Infringement Cases: A Critique of the Hearsay Rule, 90 TRADEMARK REP. 746, 747 (2000) (“survey evidence has become de rigueur in trademark infringement cases”; “many courts will draw an adverse inference against a plaintiff ... if a survey is not introduced.”); 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) (surveys are “all but indispensable” in trademark litigation); Vision Sports, Inc. v. Melville Corp., 888 F.2d 609, 615 (9th Cir. 1989) (“An expert survey of purchasers can provide the most persuasive evidence of secondary meaning.”)

into the interpretation of contract language. Contract promises, like marketing communications, are messages transmitted by one party to another. When a firm drafts a promise (or a disclaimer) into the standard contract terms, it is communicating with the same customers to whom advertising and branding statements are made. If the survey method is a reliable method to elicit the meaning of the informal marketing-phase messages, why isn’t it equally instructive for deciphering the meaning of drafted contractual terms? We therefore begin this section in part II.A by reviewing the survey method in trademark and unfair competition law. Part II.B then presents the survey method in contracts disputes.

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 the mark, which consists of generic terms, has acquired secondary meaning. And conversely, trademark protection may be lost if a mark that consists of fanciful terms has become generic. Surveys are also dispositive when a court must decide whether a mark has been infringed. In all these inquires, the question is whether a likelihood of consumer confusion exists. Surveys address this question directly, asking people to report the inferences they make when seeing the mark.

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.” Although the Restatement on Unfair Competition counsels against this approach, “many courts will draw an adverse inference against a plaintiff ... if a survey is not introduced.”

39 MARY LAFRANCE, UNDERSTANDING TRADEMARK LAW 59 (2005).

40 See Diamond & Franklyn, supra note 38, at 2032-40.

41 Id.


Survey evidence is regularly dispositive at the summary judgment stage. Courts often grant summary judgment based on survey evidence and they overrule summary judgment when survey evidence shows an issue of material fact. Defendants can succeed in offering exculpatory surveys, 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.

Commentators document the central role of survey evidence in trademark litigation based on various methodologies, particularly by looking at reported case outcomes. Looking only at published opinions may create a biased impression of the impact of surveys. But 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,” playing “key roles in claim evaluation and are understood by attorneys as an influential settlement tool for both sides.” 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.

Survey evidence has similar crucial applications in other areas of unfair competition law, and in particular in false advertising claims where the allegation is that consumers were misled or confused. For example, an allegation that the defendant’s

---


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

47 Florida Int’l Univ. Bd. of Trustees v. Florida Nat’l Univ., Inc., No. 15-11509, 2016 WL 4010164, at *16 (11th Cir. July 26, 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.”).

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


51 Diamond & Franklyn, supra note 38, at 2062.

52 Id. at 2049.
advertisement implied falsity requires evidence about the mental impressions the ad left on consumers. A leading court established that “the success of a plaintiff’s implied falsity claim usually turns on the persuasiveness of a consumer survey.”

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. 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. Judges exclude surveys that fail to satisfy such basic reliability standards under Federal Rule of Evidence 403. Even if admitted, the weight of survey evidence varies according to, among other factors, how well the survey captured the relevant “universe” of consumers, how representative the sample from that universe was, whether a control group was maintained to exclude background noise, and whether the results of the survey were verified. Courts examine whether the design,

---

53 Johnson & Johnson*Merck Consumer Pharms. Co. v. SmithKline Beecham Corp., 960 F.2d 294 (2d. Cir. 1992). See also Novartis Consumer Health, Inc. v. Johnson & Johnson-Merck Consumer Pharms. 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.”).

54 Laura Heymann, Surveying the Field: The Role of Surveys in Trademark Litigation, JOTWELL (February 23, 2015) 342.


56 See Thornburg, supra note 44, at 93. 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”).

57 McCARTHY, supra note 37, 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.”)


60 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); see also Toys “R” Us, Inc. v Canarsie Kiddie Shop, Inc.,
questionnaires, and interviews were unbiased, and scrutinize the accuracy of the data analysis and report.61

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.62 But courts that take a more aggressive approach towards survey results that are more prejudicial than probative, excluding such evidence.63 Over time, courts have managed to reach greater convergence over the appropriate trademark survey methodologies. The Federal Judicial Center publishes a reference guide on survey research that was authored by one of the field’s leading academics.64 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.”65 Few scholarly commentators doubt the efficacy of survey evidence per se, with the consensus being that “the field of survey research incorporates all the essential structural techniques of other scientific expert evidence,

559 F Supp 1189, 1205 (EDNY 1983) for summary of the Second Circuit’s requirements for surveys.

61 J & J Snack Foods, Corp. v. Earthgrains Co., 220 F. Supp. 2d 358, 369 (D.N.J. 2002); see also Exxon Corp. v. Xoil Energy Res., Inc., 552 F. Supp. 1008, 1022 (S.D.N.Y. 1981) (“The evidentiary value of plaintiffs survey is lessened considerably by the absence of practices and procedures which courts have found useful in assessing the validity of survey results in trademark infringement cases, including: use of only non-leading questions; verification by re-interviewing a substantial number of those interviewed, design by qualified experts; and administration by impartial interviewers.”). See generally McCARTHY, supra note 37, at § 32:171 (4th ed.).

62 MasterCard Int’l Inc. v. First Nat. Bank of Omaha, Inc., 2004 WL 326708, at 7 (S.D.N.Y. 2004). See also Thornburg, supra note 44, at 718 (noting that “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.”).


including rigorous hypothesis testing, experimental design, control conditions, and statistical inference."\(^{66}\)

There are occasional dissenting voices, and it might not surprise anyone that one of them came from Judge Posner, who referred to statistical surveys as “a black art.”\(^{67}\) Writing in 1994, Judge Posner seemed to deny that 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 fool-proof.”\(^{68}\) 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.\(^{69}\) 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 condemn 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.\(^{70}\) 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.”\(^{71}\) 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.”\(^{72}\)

Trademark surveys are certainly imperfect, but the consensus is that they are preferable on balance to alternative methods for establishing the meaning of marketing messages and the likelihood of consumer confusion. A quick examination of the caselaw suggests that the survey skeptics’ arguments have not carried the day. Surveys that


\(^{67}\) Indianapolis Colts, Inc. v. Metropolitan Baltimore Football Club Limited Partnership, 34 F3d 410, 416 (7th Cir. 1994).


\(^{72}\) Muha v. Encore Receivable Mgmt. Inc., 558 F.3d 623, 628 (7th Cir. 2009).
speak to likelihood of confusion use several standardized formats,73 approved by federal appellate courts.74 Similarly, surveys establishing whether a mark is generic are subject to court-imposed methodological guidelines.75 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.76 Surveys can be manipulated so as to elicit the desired results; they also can be resampled or selectively presented.77 Even academic surveys are prone to this defect, presenting difficulty for the research 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.78 The solutions typically applied to such credibility challenges could provide some assurance, and we discuss them in section 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


74 For a framework set by the Seventh Circuit, see Union Carbide Corp. v. Ever-Ready, Inc. 531 F.2d 366 (7th Cir. 1976) (providing a framework for presentation of survey questions). See also Simon Prop. Grp. L.P. v. mySimon, Inc., 104 F. Supp. 2d 1033, 1050-51 (S.D. Ind. 2000) (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 (8th Cir, 1980).

75 220 F Supp 2d 358, 369-372 (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).


77 Kraft Foods Grp. Brands LLC v. Cracker Barrel Old Country Store, Inc., 735 F.3d 735, 741 (7th Cir. 2013). (“[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 fool-proof.”)

78 See Swann, supra note 70, at 920-22.
fairly representative picture,"\(^\text{79}\) or engage in “an exercise in pure judicial fantasy”\(^\text{80}\) 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 increasingly explicit in conceding that without survey evidence they are unlikely to reach proper conclusions.\(^\text{81}\) Trademark and unfair competition law ask a variety of empirical questions and is increasingly using empirical tools to answer them.

**B. The Survey Method in Contract Interpretation**

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 them instead to rely on what large samples of respondents, who resemble the contract’s audience, say.

If the survey produces a winner—an interpretation supported by a proper 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. Indeed, under the survey interpretation method the winner ought to prevail on a motion for summary judgment.

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


\(^{80}\) Sun life Assurance Co. of Canada v. Sunlife Juice Ltd. (1988), 22 C.P.R. (3d) 244, 249 (Ont. Sup. Ct.).

\(^{81}\) For a particularly candid assessment, see Judge Weinstein’s opinion in **Verizon Directories Corp. v. Yellow Book USA**, Inc., 309 F. Supp.2d 401, 407 (E.D.N.Y. 2004) (“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.’ Yet, that is what the court is expected to do when asked to characterize a video as ‘mere puffery.’ Arguably, the communication intended by [defendant] and understood by the viewer is defendant’s … It may, however, 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.”).
Williston/Corbin divide on text v. context. 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 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 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. It is not the only challenge. Another set of questions is whether respondents have to be shown only the plain text and the two proposed interpretations, or should they 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. There is no demonstrably superior approach here. 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.

A host of technical questions also arise 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 statistically significantly different from a 50-50 split? Should courts count differently respondents who say that the language “definitely” means X versus those who say that are less decisive? We think that 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.

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. This is a very low threshold, suitable

82 See, e.g., ROBERT A. HILLMAN, THE PRINCIPLES OF CONTRACT LAW 269-71 (2004); FARNSWORTH, supra note 10, at 422.

83 Compare 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.”).

perhaps to the protection of a proprietary mark, but not to the interpretation of contractual language. As we will discuss below, a sizeable minority of even well-compensated survey respondents are going to answer at random when faced with questions whose answers they haven’t 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-proferentum interpretation (applicable primarily in insurance contract law), courts are instructed to interpret ambiguous terms against the drafter. This principle can be implemented within the survey method by requiring the drafting party to achieve a “super majority” 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. 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 language in contracts, a term that fails this standard could be “sanctioned” by having to achieve a super majority.

An implicit question in applying the presumption against the drafter (or other interpretive presumptions) is how strong should they be? In the context of the survey method, the question is what majority support would the drafter 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.

So far, our discussion of the survey interpretation method envisioned a consumer contract, drafted unilaterally by a sophisticated party. But the 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 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 (especially if many “insiders” have a

85 See infra section III.B.3.

86 See 5 MARGARET N. KNIFFIN, CORBIN ON CONTRACTS: INTERPRETATION OF CONTRACTS § 24.27 (Joseph M. Perillo ed., rev. ed. 1998) (explaining that contra proferentum is “technique” in which courts “adopt the meaning that is less favorable in its legal effect to the party who chose the words”); Gaunt v. John Hancock Mut. Life Ins. Co., 160 F.2d 599, 601–02 (2d Cir. 1947) (“[C]ontra proferentum 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.”)
stake in how the interpretive battle is resolved), and the size of the survey might have to be smaller. Here, however, a simple majority would suffice for one party’s interpretation to prevail (unless there is a large undecided group).

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 survey evidence produced by the other side and to highlight any defects. The survey method is readily amenable to experimental interventions. Subjects can be randomly assigned to different treatments (e.g., prompts, sequence of questions, phraseology) 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. (In Part III we demonstrate how this advantage provides insights not otherwise available).

Second, courts could threaten to completely disregard surveys that are tainted, in order to induce parties to provide competent ones. 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. The benefit to each party of designing a biased survey is at least partially offset by risk of having the survey “defeated” by the other party’s more reasonable design.

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.


90 See, e.g., Calvin Klein Cosmetics Corp. v. Parfums de Coeur Ltd., 824 F.2d 665, 670-72 (8th Cir. 1987) (approving trial courts appointment of a special master to evaluate survey and other evidence in a trademark dispute); Mallatier, 525 F. Supp. 2d. at 563 (in which the court followed special master academics’ recommendations to exclude several expert surveys on the basis of methodological flaws); SunAmerica Corp. v. Sun Life Assurance Co. of Canada, 890 F. Supp. 1559, 1570-71 (N.D. Ga. 1994) (joint survey conducted by the opposing litigants, at the court’s suggestion), vacated on other grounds 77 F.3d 1325 (11th Cir. 1996).
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.\textsuperscript{91} The procedural authority to do so, of course, exists.\textsuperscript{92} In anticipation, the
parties may agree that the court would select the survey firm. Or, the parties may
contract into ADR 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 that 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
help parties ascertain the meaning also at the time of drafting the contract. 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, Parts IV and V will develop the fuller normative (part IV) and doctrinal
(part V) cases 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.

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

\textsuperscript{91} Phyllis J. Welter, \textit{A Call to Improve Trademark Survey Evidence}, 85 \textit{TRADEMARK REP.} 205, 209
(1995). See also \textit{Indianapolis Colts}, 34 F.3d at 414 (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.”)

\textsuperscript{92} Fed. R. Evid. 706
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 dispute 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:

- 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 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. One advantage of the Toluna sample is that it is used relatively rarely by academic researchers, rendering the sample less polluted for our purposes. Wave 1 of the sample surveyed 1300 respondents and

---

93 We randomized the order -- whether the continuum began with the plaintiff or with the defendant being definitely correct. The randomly assigned order was held consistent throughout the survey for each respondent.

94 In wave 1, the mean age of respondents was 45.7 with a range of 18-90 and a standard deviation of 16.1. Females comprised 51.0% of the sample. 81.5% of the sample 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
Wave 2 surveyed 1294 respondents. In each wave, a few responses were discarded based on having completed 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.\footnote{For a discussion of the demographics of non-internet users in the United States, see Andrew Perrin & Maeve Duggan, Americans’ Internet Access 2000-2015, Pew Research Center, June 26, 2015, available at http://www.pewinternet.org/2015/06/26/americans-internet-access-2000-2015/ .}

B. Results

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.

We chose a term in insurance contracts that has long split courts, and which has often been characterized by courts as ambiguous. (Many courts view it as unambiguous, but they are almost evenly split as to which way.)\footnote{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, on the question of liability for babysitting injuries when a sitter was also caring for a relative).} The term deals, not surprisingly, with coverage exclusion (often phrased as double- or triple-negative). Specifically, standard homeowners’ insurance policies contain coverage against tort liability arising from injuries to house visitors. 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,\footnote{430 N.E.2d 641 (Ill. App. 1981).} in which the homeowner, Rebecca Moore, was watching 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 18-87 and a standard deviation of 16.94. Females comprised 52.6% of the sample. 79.2% of the sample 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. 11% 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.
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.”

The trial court in Illinois granted summary judgment in State Farm’s favor. It interpreted the language to unambiguously exclude the injuries. The appellate court reversed. Focusing on the phrase “except activities ... ordinarily incident to nonbusiness pursuits,”98 two judges held that his exception to the exclusion applied unambiguously (because boiling a pot of water was something Moore would have done for non-business reasons). They also cited the many prior cases that interpreted this exact phrase and found that courts reached conflicting holdings. This suggested to the majority judges that at the very least the language was ambiguous in its application to situations like Moore’s.99 And when insurance contract language is ambiguous, courts interpret it against the insurer. The dissenting judge sided with the trial court. He opined that the exclusion applied unambiguously because the water was boiled for a meal prepared for the neighbors’ child, for which Moore was being paid as part of her licensed business.100

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

<table>
<thead>
<tr>
<th>Injury DEFINITELY covered</th>
<th>Frequency</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>264</td>
<td>20.3</td>
</tr>
<tr>
<td>Injury PROBABLY covered</td>
<td>280</td>
<td>21.6</td>
</tr>
<tr>
<td>Coverage is COMPLETELY UNCERTAIN</td>
<td>231</td>
<td>17.8</td>
</tr>
<tr>
<td>Injury PROBABLY NOT covered</td>
<td>307</td>
<td>23.7</td>
</tr>
<tr>
<td>Injury DEFINITELY NOT covered</td>
<td>216</td>
<td>16.6</td>
</tr>
<tr>
<td>Total</td>
<td>1298</td>
<td>100.0</td>
</tr>
</tbody>
</table>

Respondents split almost evenly between the homeowner’s and the insurance company’s interpretations of the policy. About 42% said that the policy either definitely

98 Id. at 646-47.
99 Id. at 647.
100 Id. at 648 (Reinhard, J., dissenting).
or probably covered the injuries and about 40% said that the policy definitely or probably did not cover the injuries. Only 37% of respondents said the answer was definite in either direction. This is what ambiguity looks like, with our respondents following a similar equal split as the judges in the Moore case and in other insurance coverage cases.

Given this assuring baseline, 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 converge on a non-ambiguous interpretation. To that end, in Wave 2 we split our respondents. Half of them (the control group) were asked exactly the same question as in Wave 1. The other half (the treatment group) were shown shorter, seemingly less ambiguous, contractual language, one that reads:

“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 non-business pursuits” we expected fewer respondents would choose “It’s completely uncertain.” Second, without the exception to the exclusion, the exclusion could only be broadened, and there we expected more respondents to side with the insurance company. We received the following results:

<table>
<thead>
<tr>
<th></th>
<th>Original Exclusion</th>
<th>Shortened Exclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Frequency</td>
<td>Percent</td>
</tr>
<tr>
<td>Injury DEFINITELY covered</td>
<td>98</td>
<td>15.2</td>
</tr>
<tr>
<td>Injury PROBABLY covered</td>
<td>121</td>
<td>18.8</td>
</tr>
<tr>
<td>UNCERTAIN</td>
<td>122</td>
<td>18.9</td>
</tr>
<tr>
<td>Injury PROBABLY NOT covered</td>
<td>153</td>
<td>23.8</td>
</tr>
<tr>
<td>Injury DEFINITELY NOT covered</td>
<td>150</td>
<td>23.3</td>
</tr>
<tr>
<td>Total</td>
<td>644</td>
<td>100.0</td>
</tr>
</tbody>
</table>

Sure enough, respondent 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, compared to 28% who sided with the policyholder).\textsuperscript{110} The percentage of the sample saying the policy’s meaning was

\textsuperscript{110} To our surprise, the control group’s behavior did not mimic that of the Wave 1 respondents, siding more with the insurer (47% to 34%). This failure of replication did not recur in our other experiments.
uncertain also declined (from 19% to 15%). This tells us that respondents as a whole reliably pick up 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 to embrace the company’s preferred interpretation.

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

2. Ambiguous bonus agreement

The first survey dealt with a classic case of ambiguous language. A long line of cases and a simple reading of the term suggested ambiguity, and we saw that respondents reached the same conclusion. In our second survey, we wanted to see how respondents decide a case that judges viewed as unambiguous, but where different judges assigned to it a different (unambiguous) meaning. 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. Such was the issue in an Illinois Supreme Court case, Storybrook v. Carlson, in which the employer and its two employees agreed on a profit sharing bonus that stated:

The bonus shall be computed in the following manner:

- 0 to $10,000 Net Profit: No Bonus shall be paid to each employee
- $10,000 to $20,000: a maximum bonus of 5% shall be paid to each employee
- $20,000 and over: a maximum bonus of 22% shall be paid to each employee.

Subsequently, the firm earned profits exceeding $20,000. The question came up: are the employees entitled to 22% of all profits (employees’ position) or only 22% of those profits above $20,000, and a smaller percentage (0% and 5%) of the rest (employer’s)? The question was submitted to the jury, which sided with the employees. 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 employers’ 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

---


103 312 N.E.2d 27 (Ill. 1974).
to the last clause in the excerpt – to support the conclusion.\textsuperscript{104} The Illinois Supreme Court affirmed the trial court’s judgment unanimously on both grounds, saying that language at issue “so overwhelmingly” favored employer’s interpretation that “no contrary verdict … could ever stand.”\textsuperscript{105}

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 where asked to pick between the employee’s claim of $5000 bonus (20\% of the entire profit) and the employer interpretation of $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 results to the question of “How much does the employer owe the employee?” were:

<table>
<thead>
<tr>
<th>Table 2A: Bonus Agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency</td>
</tr>
<tr>
<td>DEFINITELY $5,000</td>
</tr>
<tr>
<td>PROBABLY $5,000</td>
</tr>
<tr>
<td>It is COMPLETELY UNCERTAIN whether it is appropriate to pay the employee $5,000 or $2,000</td>
</tr>
<tr>
<td>PROBABLY $2,000</td>
</tr>
<tr>
<td>DEFINITELY $2,000</td>
</tr>
<tr>
<td>Total</td>
</tr>
</tbody>
</table>

Interestingly, the respondents had a reaction similar to the jury’s— the employee’s (more generous) interpretation of the contract language was the more reasonable interpretation. About 49\% of respondents in Wave 1 thought the employee’s interpretation was definitely or probably correct, versus 28\% who thought the

\textsuperscript{104} Id. at 29. The trial judge also held that under the employees’ interpretation a few dollars’ difference could have major economic consequences. It was also unlikely the firm intended to award even two important employees 22\% of their profits, each.

\textsuperscript{105} Id. at 29-30.
employer’s interpretation was definitely or probably correct. Among those who took a position, almost two-thirds sided with the employee’s interpretation. A lot of respondents thought the language completely ambiguous.

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 proferentum logic—a bias against the drafting party whose carelessness created the ambiguity. To arbitrate between the two conjectures, we reran the survey in Wave 2. This time we showed half the group the same language as in Wave 1, and the other half a clearer text:

“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>
<thead>
<tr>
<th></th>
<th>Original Bonus Agreement</th>
<th>Clearer (Pro-Employer) Bonus Agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Frequency</td>
<td>Percent</td>
</tr>
<tr>
<td>DEFINITELY $5,000</td>
<td>181</td>
<td>28.5</td>
</tr>
<tr>
<td>PROBABLY $5,000</td>
<td>139</td>
<td>21.9</td>
</tr>
<tr>
<td>It is COMPLETELY UNCERTAIN whether it is appropriate to pay the employee $5,000 or $2,000</td>
<td>134</td>
<td>21.1</td>
</tr>
<tr>
<td>PROBABLY $2,000</td>
<td>107</td>
<td>16.9</td>
</tr>
<tr>
<td>DEFINITELY $2,000</td>
<td>74</td>
<td>11.7</td>
</tr>
<tr>
<td>Total</td>
<td>635</td>
<td>100.0</td>
</tr>
</tbody>
</table>

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. The percentage of respondents saying the employer’s position was definitely correct nearly 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 serious
mathematical deficiencies. But the strong majority should suffice to help courts choose the pro-employer interpretation.\textsuperscript{106}

In sum, experiment \#2 provides ample reason to critique what judges did in \textit{Storybrook}. The state supreme court’s determination that the bonus provision “overwhelmingly favored” \textit{Storybrook}’s interpretation was not an accurate reflection of how employees, perhaps handicapped by their less than excellent language artistry, understand 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, \textit{Storybrook} could have made small tweaks to its employment agreement that would have caused the bonus term to clearly convey the drafter’s intent.

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 \textit{Vargas v. Insurance Company of North America},\textsuperscript{107} a Second Circuit opinion interpreting the language of an aviation insurance policy. 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.”\textsuperscript{108} The crash of the small airplane occurred as the plane was flying from New York to Puerto Rico, twenty five miles west of Puerto Rico, in international waters. 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

\textsuperscript{106} An alternative approach in a case like \textit{Storybook} would be to use the survey evidence to calculate a weighted average, moving the law away from a binary choice. For example, the law could fully weight all the “Definitely $5,000” and “Definitely $2,000” votes while giving 50% weight to each of the “Probably” votes and ignoring the “completely uncertain” votes. Some quick math from this method yields a mean bonus amount of $3988.10. The weighted average approach permits researchers to utilize some of the granularity that arises from a 5-point scale rather than a 2-point or 3-point scale.

\textsuperscript{107} 651 F.2d 838 (2d Cir. 1981).

\textsuperscript{108} \textit{Id.} at 839.
against the drafter.\textsuperscript{109} We showed respondents in wave 1 a vignette based on \textit{Vargas}, and obtained these results:

\begin{center}
\textbf{Table 3A: Puerto Rico Airplane Crash Insurance}
\end{center}

<table>
<thead>
<tr>
<th>Frequency</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crash DEFINITELY covered</td>
<td>240</td>
</tr>
<tr>
<td>Crash PROBABLY covered</td>
<td>199</td>
</tr>
<tr>
<td>It is COMPLETEY UNCERTAIN whether the crash is covered</td>
<td>266</td>
</tr>
<tr>
<td>Crash PROBABLY NOT covered</td>
<td>354</td>
</tr>
<tr>
<td>Crash DEFINITELY NOT covered</td>
<td>239</td>
</tr>
<tr>
<td>Total</td>
<td>1298</td>
</tr>
</tbody>
</table>

A clear plurality favored the insurer’s interpretation over the insured’s (46\% to 34\%), consistent with the trial court’s position. That said, such modest margin of prevalence, along with the high rate of the “completely uncertain” responses, could justify the appellate court’s determination that the language employed was ambiguous, such that the ambiguity should be held against the insurer. Under the survey interpretation method \textit{Vargas} is a borderline case, where deciding what to do about the 21\% of the population that viewed the policy as completely ambiguous could be outcome-determinative.

Our best judgment about how to proceed in a case like this is that the “completely uncertain” responses should, at least in part, be treated as informative abstentions. First, and notably, a majority (54\%) of the sample did not embrace the insurer’s interpretation. Second, among those supporting one of the two interpretations, the majority of 57.5\% of the respondents that prefer the insurer’s interpretation to the insured’s is probably too slim to overcome contra proferentum, especially if (as we showed in experiment \#2) clearer language could have been used. On this analysis, \textit{Vargas} was decided correctly.

To further justify the holding against the insurer, we needed to show that insurers can win interpretation battles with more convincing majorities. We already showed, in experiments \#1 and \#2, that a careful redrafting of the term could create such majorities. Here, we took a different methodological path—a comparison to a different, un-tweaked, term. In Wave 2, we chose instead a contested aviation insurance term from a different case, \textit{Security Insurance Company of Hartford v. Andersen}.\textsuperscript{110} In that case, the policy contained a condition for coverage requiring that the pilot hold a valid

\textsuperscript{109} \textit{Id.} at 840-42.

\textsuperscript{110} 763 P.2d 246 (Ariz. 1988).
and current medical certificate attesting fitness to fly. 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.\textsuperscript{111} 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.\textsuperscript{112} The Arizona Supreme Court unanimously disagreed, finding
that the insurance policy language at issue was “completely unambiguous.”\textsuperscript{113} 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 scene with debris and smoke.]

We showed respondents a vignette based on \textit{Andersen} in Wave 2 of the survey. 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”).\textsuperscript{115} 75% of those siding with one of the two interpretations agreed with the insurer, a meaningfully greater percentage than we saw in the \textit{Vargas} vignette. This level of consensus would be adequate to overcome contra proferentum, resulting in a win for the insurer.

\begin{itemize}
\item \textsuperscript{111} \textit{Id.} at 248.
\item \textsuperscript{112} \textit{Id.} at 249.
\item \textsuperscript{113} \textit{Id.}
\item \textsuperscript{114} \textit{Id.} at 250-51.
\item \textsuperscript{115} Wave 2 respondents were randomly assigned to two different conditions reflecting legally irrelevant alterations to the vignette. The control group saw 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.
\end{itemize}
Table 3B: Arizona Airplane Crash Insurance

<table>
<thead>
<tr>
<th>Crash State</th>
<th>Frequency</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crash DEFINITELY covered</td>
<td>146</td>
<td>11.3</td>
</tr>
<tr>
<td>Crash PROBABLY covered</td>
<td>137</td>
<td>10.6</td>
</tr>
<tr>
<td>It is COMPLETEY UNCERTAIN whether the crash is covered</td>
<td>168</td>
<td>13.0</td>
</tr>
<tr>
<td>Crash PROBABLY NOT covered</td>
<td>353</td>
<td>27.3</td>
</tr>
<tr>
<td>Crash DEFINITELY NOT covered</td>
<td>489</td>
<td>37.8</td>
</tr>
<tr>
<td>Total</td>
<td>1293</td>
<td>100.0</td>
</tr>
</tbody>
</table>

Under any workable legal regime, the answer provided by the Arizona Supreme Court in *Andersen* has to be the correct one, given how clear the contractual provision is. 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 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 appear to be answering at random, misunderstand the language and survey questions, or focus 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 again should be the new “100” and majority rates should be normalized in relation to it.\(^{116}\) Perhaps there should also be a sliding scale: the larger the “uncertain” block, the greater the majority required among the rest.

\(^{116}\) For example, under these assumptions, survey results of unambiguous pro-X term would be, on average, 60% to 20% in favor of X among those voting for one of the two interpretations, and 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.
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 legal in some jurisdictions, and we think there is a widespread sentiment to limit their application, especially when the employee is fired.

In *Cambridge Engineering v. Mercury Partners*,\(^\text{117}\) an employee of company A who worked as a sales representative signed a covenant not to compete. 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.\(^\text{118}\)

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, even if the employee were to take a job as a janitor for company B it would be prohibited. The second possible interpretation is narrow: the prohibition applies only to jobs with company B that directly compete with Company A.\(^\text{119}\) The court chose the broad interpretation, relying on an interpretive canon—the so called “rule against surplusage.”\(^\text{120}\)

---

\(^{117}\) 879 N.E.2d 512 (Ill. App. 2007).

\(^{118}\) Id. at 517.

\(^{119}\) Ironically, in *Cambridge* it was the employee who argued for the broad interpretation. 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. Conversely, it was the company that advocated the narrow interpretation, hoping that it would survive the enforceability test.

\(^{120}\) Id. at 517. The court emphasized that under Cambridge’s interpretation the language “for or on behalf of Employer’s competitors,” would be rendered superfluous. As a result, company A’s non-compete clause was broader than necessary and unenforceable as a matter of law.
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 company B. Identical language was tested in both waves of the survey. Here are the results:

**Table 4: Bonus Agreement**

<table>
<thead>
<tr>
<th></th>
<th>Wave 1 Frequency</th>
<th>Wave 1 Percent</th>
<th>Wave 2 Frequency</th>
<th>Wave 2 Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEFINITELY prohibited</td>
<td>424</td>
<td>32.6</td>
<td>501</td>
<td>38.7</td>
</tr>
<tr>
<td>PROBABLY prohibited</td>
<td>382</td>
<td>29.4</td>
<td>295</td>
<td>22.8</td>
</tr>
<tr>
<td>COMPLETELY UNCERTAIN</td>
<td>305</td>
<td>23.5</td>
<td>241</td>
<td>18.6</td>
</tr>
<tr>
<td>PROBABLY NOT prohibited</td>
<td>119</td>
<td>9.2</td>
<td>144</td>
<td>11.1</td>
</tr>
<tr>
<td>DEFINITELY NOT prohibited</td>
<td>69</td>
<td>5.3</td>
<td>111</td>
<td>8.6</td>
</tr>
<tr>
<td>Total</td>
<td>1299</td>
<td>100.0</td>
<td>1292</td>
<td>100.0</td>
</tr>
</tbody>
</table>

Wave 1 respondents agreed lopsidedly with the broad interpretation—that the non-compete agreement prohibited the employee from taking any position with the company A’s competitor (62% 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.”

It is encouraging that the results did not vary materially over time, suggesting that firm could safely field test contract terms at early phases and accurately anticipate their subsequent legal exposure.

An alternative possibility is that some respondents harbored antipathy towards an employee seen as disloyal, rather than pro-little guy sentiment. These varied moral frames would counteract one another in our data. It is also possible that some small 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 is that the covenant as written precluded work of any type for a competitor.
5. Interpretation of Consumer Standard Form Contract

Our final illustration concerns the interpretation of a term in Gmail’s privacy policy. We included it here even though our data has been reported elsewhere because it illustrates additional promise for the survey interpretation method. The case itself is still pending, but has dragged on for years and proved costly to resolve. The case displays some of the problems with the existing court-based interpretation. Could survey respondents 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. In its defense, Google invoked a term in its privacy policy that allegedly grants it users’ 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. In denying Google’s motion to dismiss, the Federal district court held that the term in the privacy policy was ambiguous. 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, and the case was assigned to the same federal judge. Here, the judge found that Yahoo’s privacy policy was unambiguous and provided Yahoo users with notice adequate to secure their consent. 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. This was the case even though respondents regarded the email provider’s

---

122 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. S69 (2016).

123 In reality, there were three separate versions of this term, as Google has changed the text over the years.


125 In re Yahoo Mail Litigation, 7 F. Supp.3d 1016 (N.D. Cal. 2014).

126 Strahilevitz & Kugler, supra note 122, at S77-S78. The experiment differed the experiments for this article in one aspect: respondents did not have a choice of “completely ambiguous.”
conduct at issue (automated content analysis of user emails for the purposes of showing user personalized ads) as quite intrusive.

Table 5: Gmail/Yahoo Privacy Agreement

| DEFINITELY ALLOWED | 26.6 |
| PROBABLY ALLOWED   | 38.5 |
| PROBABLY NOT ALLOWED | 13.4 |
| DEFINITELY NOT ALLOWED | 21.3 |
| Total              | 99.8 |

This is disturbing. Changes in the language that a court deemed legally decisive were regarded as irrelevant by participants in a randomized experiment. The Google and Yahoo cases therefore represent further instances in which courts 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

Each of the cases that inspired the surveys and experiments above was a tough interpretive nut for courts to crack. Each case that formed the basis for a new survey 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 and in Vargas (where courts and respondents agreed the language was ambiguous) as well as in Andersen, Cambridge, and Yahoo (where courts and respondents agreed the language was unambiguous). But the Storybrook bonus case and Gmail privacy policy case are two prominent examples in which the court missed the mark, compared to the majority understanding. The Storybrook court embraced an unambiguous meaning of words that was opposite to the dominant interpretation among lay respondents (and the jury the in case). The Gmail court held that privacy policy language was too ambiguous to secure user consent to content analysis of their emails, but respondents who read the policy language overwhelmingly disagreed. We anticipated a clash between what the court did and what

127 Strahilevitz & Kugler, supra note 122, at S80-S83, obtained similar results on 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. 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.
respondents said they would do in Storybrook, but the Gmail result took us completely 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 began 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 precedents, 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.

Having shown that the survey interpretation method is feasible, we now want to argue that it is desirable. The normative defense rests on two arguments, one focusing on the costs, and the other on the benefits, of interpretation.

A. Costs of Interpretation

Interpretation of contractual text is the most common task for contract lawyers and the most frequent ground for dispute in contract law.¹²⁸ These disputes are costly, in

¹²⁸ See generally Jeffrey M. Lipshaw, Metaphors, Models, and Meaning in Contract Law, 116 PENN. ST. L. REV. 987, 989 (2012) ("[l]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
part because they require courts to determine what extrinsic evidence to consult (and then consult it), but primarily because they drain expensive lawyer time.\(^\text{129}\)

Survey interpretation is becoming increasingly cheap. In the past, a major criticism of surveys in trademark litigation was their cost.\(^\text{130}\) The International Trademark Association, for example, reported the cost of surveys to be in the $25,000-$150,000 range.\(^\text{131}\) But the rise of online panel surveys promises a significant reduction of the cost of conducting surveys.\(^\text{132}\) 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.\(^\text{133}\) Reputable online panels provide sufficient compensation to respondents to ensure serious and meaningful responses. (Unlike the respondents contacted by phone or while shopping in malls, they are not being asked to donate their time.) Online surveys have the distinct advantage of allowing respondents to examine visual stimuli and texts with adequate time to ponder what they see.\(^\text{134}\) The selection of the respondents’ panels is the real world of contracts is almost exclusively about negotiating and writing documents and perhaps interpreting them later (whether or not they get litigated”).


\(^\text{130}\) See Diamond & Franklyn, *supra* note 38, at 2054, (cost might lead “[c]lients who may benefit from surveys [to be] potentially priced out of court” and “[t]he 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.”) See also Thornburn, *supra* note &&, at 718 (“current survey experts in California charge between $450 to $600 per hour and require support staff billing at rates ranging between $200-300 in orchestrating the actual surveys”).


\(^\text{132}\) The online survey firms we used for the experiments reported in part III charge $3.25 per respondent, for a 15-20 minute survey. Telephone interview, by contrast, cost on average $20 per respondent. See Karin Braunsberger, Hans Wybenga, Roger Gates, *A comparison of reliability between telephone and web-based surveys*, *Journal of Business Research*, Volume 60, Issue 7, 758, 763-64 (July 2007).


\(^\text{134}\) 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,
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. There is the additional cost of designing the survey in a professional manner and persuading the court that it is done reliably. That said, the techniques necessary are not exactly rocket science.

Low survey costs make the method more attractive, but potentially also more manipulable. As the costs of conducting online surveys declines, the danger that a party will try out multiple versions of surveys, cherry-pick the most favorable results, and bury the less favorable ones. It is believed that charlatans, “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,” operate in the trademark litigation landscape. As we discussed in Part II above, the use of neutral experts chosen either by court or in the contract would help address this problem, as would rules requiring parties to disclose in litigation all preliminary and final conclusions reached by experts whose testimony parties elect not to introduce in litigation. One nice aspect of the survey method under the adversarial system is that it promotes immediate replication, and experts whose findings were consistently outliers might suffer reputational damage.

B. Benefits of Survey Interpretation

The survey interpretation may be cheaper than conventional methods of resolving interpretation disputes. But cheap cannot be the only goal or else a coin-toss—would be superior. The question we now want to address is the benefit side of the survey interpretation method. Does it deliver results that are aligned with acceptable conceptions of accuracy and with the goals of inducing good primary behavior among contracting parties?

An important benefit of the survey method is to align the legal interpretation with the meaning that the intended audience of the contract—the people whose behavior it governs—assigns to it. 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 understanding. A commercial contract is written for merchants,

since inflection can alter meaning and some written statements are easier to understand when read than when heard.”)

135 Poret, supra note 133, at 757.

136 Thornburg, supra note 44, at 97; see also McCarthy, supra note 37, at §32:179 (discussing protection of surveys from discovery); Diamond & Franklyn, supra note 38, at 2059 (finding evidence of attorneys who collected surveys and buried those with bad results).


and its meaning should be determined by surveying people in the trade. And a sophisticated drafted-from-scratch contract is written by and for the attorneys on each side, and thus like professionals should be polled to determine the meaning.

Let us focus on mass contracts that govern transactions with lay people. 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 answer is disappointingly simple: it’s the law! The meaning even of the fine print ought to be that which a typical party to whom it is addressed would assign. 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 advertisements, offers, and contractual terms. “We give words their ‘ordinary meaning’ viewing the subject of the contract ‘as the mass of mankind would view it’.”139

Accordingly, when asked to interpret a contract under the existing jurisprudence, courts recognize that they ought to uncover the meaning that the lay parties would assign to the language, and they try to put themselves in the shoes of the non-legally trained.140 But could 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.141 The survey method would help courts overcome such cognitive and information biases.

In deciphering meaning, courts often deploy technical canons of interpretation and construction. The use of some of these techniques has fortunately subsided over time, but they are still a central feature of court-centered interpretation. Some canons are intuitive and might be followed 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 precedent them),142 ejusdem generis,143 (the meaning of the word is interpreted by the company it keeps) or expressio unius est exclusio alterus144 (to express or include one thing implies the exclusion of another), are

---

139 All-Ways Logistics, Inc. v. USA Truck, Inc. 583 F.3d 511, 516 (8th Cir. 2009).


143 See, e.g., Aspen Advisors v. United Artists Theater Co., 861 A.2d 1251, 1265 (Del. 2004); i LLC v. Greater N.Y. Mut. Ins. Co., 31 A.D.3d 100, 103-104, 815 N.Y.S.2d 507 (1st Dept. 2006) (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.)

more technical and might not resonate at all with the ways in which lay people interpret contracts.\textsuperscript{145} The survey interpretation method would deny drafters of legalese the privilege of decoding their mysterious intentions through the meticulous toolkit that canons provide. This would only serve to improve these drafters’ communications skills.

Even in those rare instances where a contract interpretation case goes to a jury, we expect a divergence between the jury’s understanding and the understanding reflected in the survey data. Most obviously, the larger \( n \) will reduce the variance and enhance the validity of the group judgment as a representation of the population at large. More subtly, when jurors convene to determine the meaning of a contractual provision, they are deliberating before reaching a judgment. Such deliberation is highly artificial if the legal system is hoping to replicate the thought process of a typical consumer. Respondents in surveys and experiments do not deliberate. They are isolated in their analysis and judgments, just like consumers or employees are when provided with the opportunity to review boilerplate language. Group deliberations are known to lead towards polarization,\textsuperscript{146} and jury deliberations often accord jurors disproportionate influence based to members of the jury based on social status, education, race, and gender.\textsuperscript{147} Lay people do not enter into contracts after deliberating over the fine print with eleven other people. To the extent that they glance at the relevant contractual language at all, they are likely to do so alone.

Interpreting contractual meaning “as the mass of mankind would view it” would have several effects on the substantive criteria used to resolve interpretation disputes, denying the opportunities to choose interpretations that serve various social goals. It would sharply conflict with the view that interpretation is a normative exercise, through which courts infuse the contract with content that promotes social ends.\textsuperscript{148} For example, it would cripple courts’ ability to promote economic efficiency by choosing the meaning

\textsuperscript{145} In influential work by Gluck and Bressman has shown that even legislative staffers who draft the laws reject or are unaware of several key canons that courts use to interpret statutes. See, e.g., Abbe R. Gluck & Lisa Schultz Bressman, Statutory Interpretation from the Inside – An Empirical Study of Congressional Drafting, Delegation, and the Canons: Part One, 65 STAN. L. REV. 901 (2013). We know of no empirical research that examines the extent of lay peoples’ awareness of interpretive canons but we would be shocked if such awareness were remotely prevalent.


that minimizes transactions costs or maximizes the parties’ joint surplus. We are not aware, however, of any evidence that judges currently try or succeed to interpret contracts in a welfare-maximizing manner, or that courts have the kind of information necessary to advance this goal. In fact, many interpretation disputes involve purely distributive aspects of the transaction, which (by definition) cannot be resolved via the welfare maximizing principle. Thus, it is unlikely that the survey interpretation method would resolve disputes in a systematically less efficient manner.

To be sure, there are other goals that court may hope to advance in the course of resolving interpretation disputes, such as favoring weak parties, reduction of transactions costs, and promotion of 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 some subset of the parties, it would have to rely on carefully selected samples to elicit this meaning. Even when a prevailing meaning is verified, courts may want to advance a different meaning for good reasons. Under the survey method, such doctrines could no longer be called “interpretation” and other doctrinal levers would have to be supplied to advance such ends.

Another artifact of reducing the interpretive discretion of court would be on the impact of context. Survey respondents would likely be presented with the relevant text of the agreement and only a very thin account of the context. It would be difficult to show them lots of additional facts regarding extrinsic communications between the parties, the norms that permeate in the trade, or the course of dealings and performance surrounding this contract, without taking more of their time, compromising the care with which they read each vignette. The interpretation elicited from survey respondents would thus be more formalistic, relying primarily on the text.

Formalism in interpretation has well known downsides. It generates outcomes that are less allegiant to the full factual account underlying the case, and it is thus less accurate in vindicating the parties’ “true” intent. But it also produces important benefits, mostly in simplicity and reduced litigation cost. It is for this reason that some jurisdictions use formalist approaches to limit the contextual inquiry in interpretation (like a strict the parol evidence rule). The survey interpretation method would further cement this direction.

---

149 See, e.g., George Cohen, Interpretation and Implied Terms in Contract Law, 2 ENCYCLOPEDIA OF LAW AND ECONOMICS 125 (De Geest, ed., 2nd Ed. 2011); Posner, supra note 129; Juliet P. Kostritsky, Interpretive Risk and Contract Interpretation: A Suggested Approach for Maximizing Value, 2 ECONOMICS 125 (De Geest, ed., 2nd Ed. 2011); Compare Lipshaw, supra note 128, at 1019.


152 See Lisa Bernstein, Merchant Law in a Merchant Court: Rethinking the Code’s Search for Immanent Business Norms, 144 UNIV. PENN. L. REV. 1765 (1996).
These features of the survey method—formalism and the improved match with lay understanding—may have several desirable ex ante effects on behavior. First, they might induce firms to draft simpler and shorter contracts with less cryptic jargon to govern their relationship with consumers and lay parties. The rise of digital contracting made it all the more costless for drafters to err on the side of verbosity, writing long sentences, clauses, and contracts. Even those that try heroically to use lay language fall back on technical legalese when dealing with weighty matters. These incentives for excess and complexity would be constrained if that language has to be understood by lay readers. Drafters of agreements would have to consider how average respondents who are not trained in legal subtlety would read and understand the term in a future survey. No longer able to rely on professional judges invoking interpretive canons and reading between the lines, drafters’ best strategy would be to eliminate excess and clutter and write clearly. Various regulatory efforts prompt firms to do just that, but without giving real incentive; the survey method can create the desired incentives for simplicity.

Second, simpler and shorter contracts that mean what they say might provide significant benefits to the few that do occasionally read contractual boilerplate. To be sure, not many do. But for the occasional “readers” the stakes of the transaction are idiosyncratically high. Making it easier for this small group to understand their contracts without forcing them to retain counsel is a benefit. Who knows, simplifying contractual language and making it comprehensible to lay audiences might even boost readership. While this effect would likely be small, it would be marginally stronger if consumers understand that the meaning of a contract is based on the responses of people just like them. The benefits of simplification would be even more pronounced if consumers are assisted by information intermediaries that are not lawyers when they enter into contracts. Homebuyers and sellers are typically advised by real estate agents or brokers rather than lawyers. Consumers are advised by services that rate or grade contractual aspects like privacy policies, warranties, and termination clauses. In these situations, making the language more readily comprehensible could have benefits even for people who do not bother to read the language they are signing.

Finally, but no less important, the survey method could make the interpretation of a contract more predictable ex ante. 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

---

153 For example, the Truth in Lending Act requires “clear and conspicuous” disclosures to consumers by credit card issuers, 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 (2012).

154 See Omri Ben-Shahar & Carl Schneider, More Than You Wanted To Know: The Failure of Mandated Disclosure, Ch. 7 (2014).

155 See Porat & Strahilevitz, supra note 34, at 1450-51.
engaging in the bargained-for conduct, all in order 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. Through the use of randomized experiments it would also give them the opportunity to prevent contractual misunderstandings before they happen and to test out changes that they want to make in contractual terms. Thus, rather than await the dispute and its uncertain resolution, parties can predict the legal outcome by applying the survey method privately.

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. 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 priors. By contrast, there is evidence suggesting that survey responses to interpretive questions will be stable over time.\textsuperscript{156}

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 requires them to await court interpretation, the parties prefer to cut-and-paste existing clauses that were already interpreted by courts and have clear and predictable meaning.\textsuperscript{157} 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.\textsuperscript{158} Survey interpretation—by allowing firms to resolve the uncertainty prior to litigation—can foster greater drafting flexibility.

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 that relies on jurists to analyze language written for

\textsuperscript{156} See Matthew B. Kugler & Lior Jacob Strahilevitz, The Myth of Fourth Amendment Circularity (2017 manuscript, on file with author).


\textsuperscript{158} Boardman, supra note 25, at 179.
people 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 section 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 some 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 exception, and judicial skepticism as to whether such surveys could be accurate. Within a few decades the opposition to consumer survey evidence had been overcome. And whatever concerns judges may have initially had about the use of online surveys in trademark litigation dissolved quickly enough. So much that it is now considered an “abuse of discretion” to exclude even an online survey on the grounds that it did not “replicate real world conditions.” 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.

159 See supra text accompanying notes 1-6.

160 Elgin Nat. Watch Co. v. Elgin Clock Co., 26 F.2d 376, 377-78 (D. Del. 1928) (in which an innovative lawyer attempted to introduce an expert-conducted consumer survey to avoid the necessity of calling 20 or 30 ordinary consumers to testify whether they had been confused. The court held that the hearsay rules give it no quarter to consider such evidence, insisting that it could only accept evidence from consumers who would testify personally in open court.)

161 Du Pont Cellophane Co. v. Waxed Products Co., 6 F. Supp. 859, 884-85 (E.D.N.Y. 1934), modified but not on this ground, 85 F.2d 75, 80 (2d Cir. 1936).


164 Fortune Dynamic, Inc. v. Victoria’s Secret Stores Brand Mgmt., 618 F.3d 1025, 1036-37 (9th Cir. 2010).

165 POM Wonderful LLC v. Organic Juice USA, Inc., 769 F. Supp.2d 188, 199-201 (SDNY 2011). 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
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?\textsuperscript{166} 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.\textsuperscript{167} And indeed, so 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 their prejudicial effect.”\textsuperscript{168} 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 contract litigation is what reasonable expectations did a promise evoke? Contracting parties claim that their reasonable expectations ought to be enforced—and courts willing to enforce such expectations are increasingly relying on surveys to establish the existence and content of the expectations.\textsuperscript{169} Or, under the doctrine of unconscionability courts have to determine, among other things, whether a term is surprising and unexpected to the typical consumer. Under the American Law Institute’s proposed Restatement of Consumer Contracts, this can be shown by survey evidence.\textsuperscript{170}


\textsuperscript{169} Walker v. Life Insurance Co. of the Southwest, No. CV 10-9198 JVS (RNBx), 2015 BL 459212 (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).

Surveys are also useful in administering another contact 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. 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 relevant commerce? If parties contract over “Grade A” chicken, an “active breeder” bull, or a “sound” horse, 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 the court’s finding.

This is also the view of the statute. “The existence and scope of such a usage must be proven as facts,” says the UCC, and courts have thus explicitly endorsed and encouraged the use of proof methods akin to surveys. But not always large surveys. Instead, a common method of proof of trade usage is experts’ testimony, usually qualitative accounts by members of the trade. This is a type of survey, but of 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. To be sure, Lisa Bernstein argued that courts rarely resort to rigorous survey evidence regarding the content or scope of trade usages. But she views the infrequent use of surveys as an unfortunate failure of the courts to implement the most appropriate and probative methodology.

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

---

171 Uniform Commercial Code § 1-205
172 Uniform Commercial Code § 1-103(c).
173 See, e.g., Timeline, Inc. v. Proclarity Corp., No. C05-1013JLR, 2007 WL 1574069, at *8 (W.D. Wash. May 29, 2007), where the court implicitly endorsed the use of survey evidence in contract interpretation, criticizing an expert trade witness for failing to “identify any surveys or research he has conducted regarding the meaning or usage of these terms; see also Lion Oil Trading & Transp., Inc. v. Statoil Mktg. & Trading (US) Inc., No. 08 CIV. 11315 WHP, 2011 WL 855876, at *2 (S.D.N.Y. Feb. 28, 2011) (a “survey of crude oil traders” to determine a trade usage). Compare 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).
174 FARNSWORTH, supra note 10, at 41 (Under the UCC, “[a] party commonly shows a usage by producing expert witnesses who are familiar with the activity or place in which the usage is observed”); Travalia, NORDSTROM ON SALES & LEASES OF GOODS, para 3.14[c] at 244 (“[P]resumably expert testimony will be necessary to establish a trade usage”).
survey consumers who say they are in the market for the goods or services at issue. 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.

In merchant-to-merchant cases courts will want to hear from respondents who are experts in the community of merchants dealing with like matters. You might worry that respondents in merchant markets will provide self-serving answers in attempt to steer legal standards to their favor. Hopefully, a clever design of the survey instrument could defeat this bias. It would surely do better than the alternative methods of proving a trade usage, which all too often rely on testimony by insiders, who are often parties to the litigation.177

The trade usage example can be generalized. In the same way that merchants are polled to interpret the meaning of contracts that are written for them, lay respondents should be asked to interpret consumer contracts that are written to regulate their transactions. People in 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. We rely on average responses where we know that the parties’ own testimony about their subjective states of mind will be clouded by self-interest. Google’s terms of service intend to apply to any user of the internet, and so it is the meaning assigned by a representative sample of internet users that would correctly uncover the understanding of the contractual parties.

As a practical matter, we think the survey method is most likely to have its first successful use in a major class action suit like the Gmail litigation. 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 its first successful adoption, the method could spread, just as it did in trademark and false advertising law.

What about commercial contracts between sophisticated parties? How should long term supply contracts, merger agreements, or sovereign bond prospectuses be interpreted? We would like to survey sophisticated business people, but there are either not enough of them, or we can’t pay them enough to bother. Alternatively, we could survey consumers, but this would likely generate pure noise. Can people understand and elicit the exact meaning of, say, a pari passu clause in commercial bond contracts? (It baffles even seasoned jurists.178) Perhaps the right audience to survey is the guild of commercial lawyers. Since these sophisticated contracts are negotiated and co-drafted by lawyers, sometimes using language and legalese that even the business people do not fully understand, it would be natural to poll this intended audience of the texts—the legal profession. One may even invent a survey of . . . judges. Imagine the Delaware

177 *Id.*

178 See sources cited supra note 157.
chancery court interpreting a provision in a corporate charter by polling all judges on the court!

Interpretation doctrine sometimes tries to do more than elicit the literal or plain meaning. Courts may look for the purpose of the parties (if such common purpose indeed exists). Or, if a term may be consistent with two reasonable meanings, it is up to the court to assign “liability” for such confusion. There are various other policy goals that interpretation doctrine could hope to promote, and it is certainly true that survey respondents are not competent to account for 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 policy evaluation, the survey method can help establish it. If a survey shows that two meanings are equally plausible, the court could apply additional criteria to determine which party prevails.

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 to see courts turn to surveys as aid in the interpretation of contractual terms, so far they have not done so systematically, in part because lawyers have not been encouraging them to consider survey-related innovations. (And why would they? When has any professional guild recommended its replacement?)

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 can accord the parties.

1. Customizing Interpretation Law

Are interpretation doctrines default rules? In a 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 interpretive meaning. But such drafting would be an overwhelming task, given the endless potential gaps in contractual language. 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?

---

179 RESTATEMENT (SECOND) OF CONTRACTS § 202(1).

180 Id. at § 201(2).

Such opt 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,” playing “key roles in claim evaluation and are understood by attorneys as an influential settlement tool for both sides.”

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. For example, parties are allowed to override the default rules of interpretation by explicitly drafting an “entire agreement” or “merger” clauses instructing courts to ignore other parol agreements (that might otherwise be looked at), and courts ordinarily comply. Or, parties may draft an “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. 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.

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’ desire to be governed by the survey interpretation method. We also see no reason why the courts ought to

---


185 Uniform Commercial Code § 2-202 comment 2 (course of dealing and trade usage to be used as interpretive tools, “unless carefully negated” in the contract); Omri Ben-Shahar, The Tentative Case for Flexibility in Commercial Law, 66 U. Chi. L. Rev. (1999).
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 clauses, limited discovery, statute of limitations clauses, enhanced requirements of writing or other proof, and more.¹⁸⁶

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.¹⁸⁷ 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. 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.


¹⁸⁷ 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 clause that assigned dispute resolution jurisdictions to biased arbitrators. See Dan Slater, San Francisco Sues Provider of Arbitrators, Alleging Bias, Wall Street Journal (April 7, 2008), available at http://blogs.wsj.com/law/2008/04/07/san-francisco-sues-provider-of-arbitrators-alleging-bias/. As it was dealt in the arbitration context, the risk of corrupt interpretation clauses would have to be addressed through doctrines like fraud and unconscionability.
There is, to be sure, an advantage in waiting for the dispute to arise. Advance testing is over-inclusive, applying to clauses that may never get litigated or to contracts and transactions that may never be subject to litigation. This added cost of over-inclusive 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 a 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 and “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 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 it 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 boilerplate, offering to parties contract forms with known meaning (perhaps kept a trade secret), and could event warrant the meaning and insure against adverse legal interpretation. Competition would go a long way to 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. Extensions and Concluding Remarks

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, and there is no reason to think it ever will be.

188 See Adam B. Badawi, Interpretive Preferences and the Limits of the New Formalism, 6 BERKELEY BUSINESS L.J. 1, 3 (2009).
The law of interpretation would remain non-uniform, inconsistent, and costly. Judges would bemoan, if not ridicule, the procedures that the common law of contracts imposes on them when they set forth to interpret contracts.  

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 methodology. 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 the last pages, we offer some thoughts on their potential extension to other domains of litigation, but also a criterion that we think ought to limit the scope of the method.

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. 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 survey 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 denied class certification as a result.

The survey-based approach offers a markedly more satisfying way to tackle the commonality problem. Just because consumers read 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. A survey showing that respondents share common reactions to different versions of the

---

189 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.” See *Trident Center v. Connecticut General Life Insurance*, 847 F.2d 564 (9th Cir. 1988).

190 Federal Rule of Civil Procedure 23(b)(3) provides that class certification is only appropriate where common questions of law or fact predominate individual factors. See *Tyson Foods, Inc. v. Bouaphakeo*, 132 S. Ct. 1036, 1045 (2016).

191 See, e.g., *Sacred Heart Health Sys., Inc. v. Humana Military Healthcare Serv. Inc.*, 601 F.3d 1159, 1171-76 (8th Cir. 2010); *Broussard v. Meineke Discount Muffler Shops Inc.*, 159 F.3d 331, 340 (4th Cir. 1998).
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 the question 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.\textsuperscript{192} Or, in designing various disclosure laws, the question how people understand the disclosure is sometimes (although not always) critical to evaluating the adequacy of disclosure.\textsuperscript{193} 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 adequacy of jury instructions. Appellate courts regularly evaluate how jurors would respond to variations in jury instructions,\textsuperscript{194} and an experimental literature has emerged to study these questions,\textsuperscript{195} 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.\textsuperscript{196} Or they hold such variations to be harmless error.\textsuperscript{197} 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.\textsuperscript{198}

\textsuperscript{192} See Matthew B. Kugler & Lior Jacob Strahilevitz, \textit{Actual Expectations of Privacy, Fourth Amendment Doctrine, and Mosaic Theory}, 2015 \textit{SUP. CT. REV.} 205 (2016).

\textsuperscript{193} See, \textit{e.g.}, Know Before You Owe: Evolution of the Integrated TILA-RESPA Disclosures (Report presented to the CFPB, Kleimann Communication Group, 2012).

\textsuperscript{194} See, \textit{e.g.}, United States v. Southwell, 432 F.3d 1050, 1052-56 (9th Cir. 2009).


\textsuperscript{196} See, \textit{e.g.}, United States v. Fernandez, 722 F.3d 1, 15-27 (1st Cir. 2013).

\textsuperscript{197} See, \textit{e.g.}, Tillman v. Cook, 215 F.3d 1116, 1123-27 (10th Cir. 2000).

\textsuperscript{198} See, \textit{e.g.}, State v. Paul, 104 A.3d 1058, 1059-62 (N.H. 2014) (insisting, absurdly, that jurors told “[i]f you have a reasonable doubt as to whether the State has proved any one or more of the elements of the crime charged, you must find the defendant not guilty. However, 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” would appreciate the significance behind the choice of “must” in the first clause and “should” in the second, which is that juries have the power to nullify criminal convictions).
Finally, it is tempting to think that if the survey method could work in these contexts, it could be expanded further, to resolve other legal issues, such as statutory interpretation, applications of legal standards (e.g., 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 what a text written to govern their behavior says. 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 answer to these questions necessarily turns solely on the ordinary lay understanding of the text.