The Readability of Contracts: Big Data Analysis

Yonathan A. Arbel[**]

Cite as: Yonathan A. Arbel, *The Readability of Contracts: Big Data Analysis*, Forthcoming J. EMPIRICAL LEGAL STUD. 21:4 (Dec. 2024)

The Plain Language Movement waged a silent revolution in the last generation, passing nearly 800 laws nationwide with little public debate. The movement asserted that it could scientifically show that there is a widespread readability crisis in legal documents, particularly contracts, that are unreadable to most adults. This article presents the largest empirical analysis of these claims to date, utilizing a dataset of 2 million contracts spanning multiple decades and industries and applying machine learning techniques.

The study challenges fundamental tenets of the Plain Language Movement. Contrary to prevailing beliefs, consumer agreements have median reading scores almost indistinguishable from those of daily news articles. A critical evaluation further exposes that readability tools endorsed by the movement are shoddy and manipulable and can produce grade-level differences of up to 4.6 years for identical texts. Moreover, the movement's core belief that Americans cannot read past the level of an eighth grader is exposed as an unsubstantiated myth.

These findings fundamentally challenge the premises and effectiveness of one of the central consumer protection policies. These results call for a radical rethinking of legal access strategies, suggesting a shift from superficial readability metrics to addressing substantive issues in market dynamics and focusing on truly vulnerable populations. More broadly, this case study serves as a cautionary tale about the propagation of myths in legal scholarship and the potential for well-intentioned reform movements to divert attention and resources from more effective interventions.

---

* Rose Professor of Law, University of Alabama, School of Law. The author would like to acknowledge the many contributions of Samuel Becher, Michael Blasie, Shahar Dillbary, David Hoffman, Florencia Marotta-Wurgler, Ben McMichael, and Julian Nyarko. Additionally, the author is grateful for the valuable feedback and insights provided by participants at both the Conference on Empirical Legal Studies and the American Law & Economics Association annual meeting in 2023. Data necessary to replicate the results of this article are available upon request from the corresponding author. The dataset collected here is made publicly available at

* The dataset collected here is made publicly available at https://github.com/yonathanarbel/readability. Readers are invited to contribute their own collections of contracts to the dataset, and their contribution will be recognized in the collection. kindly mail yaarbel@ua.edu.

INTRODUCTION

It is by now a maxim that no one reads contracts—an unsettling conclusion for an area of law whose deepest ethos is assent (Ben-Shahar, 2009; Lemley, 2023). When individuals are unaware of the terms to which they agree, they are bound to enter unfavorable deals, give paper-thin consent, become exploitable by sophisticated counterparties, and find marketplace exchange unduly punishing (Radin, 2006; Schwarcz, 2011; Oman, 2016). These are concerns that cut deep and wide, implicating normative commitments to fairness, justice, autonomy, and efficiency. It is not surprising then that in their seminal article, Ian Ayres and Alan Schwartz (2014) consider the "no reading problem" as an organizing principle in consumer law.

This naturally leads to the why: What is the reason that so many people with direct financial stakes abstain from reading the terms of their agreements? Studies that inquired into it have raised many important reasons (Plaut & Bartlett, 2011; Macro International, 2007). People explain that they do not read because reading contracts of adhesion is futile when the terms are not negotiable (Hillman & Rachlinski, 2002; Epstein, 2006). They recognize that what firms do and what contracts say are not the same (Bebchuk & Posner, 2006; Arbel & Shapira, 2020; Becher & Zarsky, 2019). They also feel that the number of contracts in our lives is unmanageable and their content is boring. They contend they will read forms when problems arise and do not have to do so at the point of sale. And finally, because people ultimately trust, perhaps too much, the legal system to protect them, they do not feel the need to read contracts literally (Arbel & Becher, 2022).

One leading strand of the literature claims a major culprit is language. Too complex, full of gobbledygook, and despairingly arcane, contracts are described as the dreariest of literary genres (Martínez, Mollica, & Gibson, 2022; Garner, 2013). Such abstruse language far exceeds the grasp of the average American, who is repeatedly cited as having a reading level around the sixth to eighth grade (Government Accountability Office, 2006). And this readability gap is said to have real consequences; fixing it would "protect consumers," "aid competition," and "help consumers to know better their rights and duties" (73 Pa. Stat. Ann. §§ 2201–2212). There are important skeptics, of course, not the least of whom are two of the reporters for the new Restatement of Consumer Contracts (Marotta-Wurgler, 2011; Ben-Shahar & Schneider, 2014). Nonetheless, many scholars adamantly claim that reform is vital, that common law courts fail their duty by imposing a "duty to read" of what is effectively "unreadable," and that statutory reform is urgently needed. It is those voices that power one of the most influential consumer political movements of our time, the plain language movement (e.g., Black, 1981; Benoliel & Becher, 2019; Friman, 1995; Blasie, 2023).

One will be excused if they have not heard of the movement. Although it occasionally features in law reviews, its public heyday was in the 1980s. Today most of its advocacy is conducted behind the scenes by think-tanks and technocratic "centers" (Center for Plain Language, n.d.; Center for Civic Design, n.d.; UAMS Center for Health Literacy, n.d.). It would be a mistake, however, to conclude that what the movement lacks in visibility it lacks in political capital. Its widespread impact can be easily gleaned when considering the Truth in Lending Act, the UCC regulation of warranty disclaimers, the Magnuson-Moss Warranty Act (Truth in Lending Act, 2018; Magnuson-Moss Warranty Act, 1975), a growing number of state laws requiring all consumer contracts be plainly written (Pa. Stat. Ann., 2006; Minn. Stat. Ann. § 325; Conn. Gen. Stat. Ann., 2012), or a very recent Consumer Financial Protection Bureau (CFPB) rule that holds complex language an abusive trade practice (Consumer Financial Protection Bureau, 2023). The entire insurance industry was transformed by strict requirements with exacting plain language demands, often framed in specific numerical scores on readability metrics (Schwarcz, 2014; Rustad, 2022). All in all, there are nearly 800 plain language laws across the United States (Blasie, 2022). By volume alone, the plain language movement is leading the single largest consumer protection revolution of our generation. Hence, David Hoffman's recent puzzlement: "What if we'd already revolutionized contract law but no one knew it?" (Hoffman, 2022).

This Article uses the power of big data analytics and machine learning to assess the central tenets of the Plain Language Movement. It evaluates four questions that are core to the Movement's claims: (1) How do consumer contracts compare in readability to texts commonly read by American adults? (2) What is the trajectory of contract readability over recent decades? (3) How reliable are standard readability metrics when applied to legal documents? (4) Does empirical evidence support the movement's claims about American adult literacy levels?

In evaluating these questions, the Article makes four novel contributions to the literature. First, it provides the largest scale empirical analysis of contracts to date; whereas the previous generation of studies used samples of a few hundred contracts (Marotta-Wurgler & Taylor, 2013; Rustad & Koenig, 2014; Benoliel & Becher, 2019; Samples et al., 2024), this study collects 2 million contracts. Second, it introduces a novel benchmarking approach, comparing contract readability to that of texts American adults routinely consume. It shows, on this basis, that the readability of consumer contracts proves not significantly different from any other text American adults read daily. Third, it offers a critical evaluation of commonly used readability metrics, exposing their weakness and manipulability. And lastly, it challenges long-held assumptions about American adult literacy. The persistent idea that Americans cannot read past the sixth or even eighth grade is revealed to be a widespread myth. This myth grounds policy in areas well beyond consumer law, such as education and design of medical disclosures (S.C. Code Ann. § 37-4-105(B), 2015).

These results carry direct and profound policy implications for consumer protection law, calling into question the regulatory apparatus built

around plain language reform. While many contracts would benefit from clarity, the notion of a platonic ideal of plain language, universally accessible and mathematically provable, should be dispelled. Plain language reforms, though well-intentioned, come with significant costs. They are used as a political cudgel (Selsky & Komenda, 2023), they divert regulatory attention from substantive issues to drafting (Oman, 2016; Zamir & Ayres, 2020; Lemley, 2023), compliance, and auditing of hundreds of rules. This approach inadvertently enshrines non-reading as a moral failure (Wilkinson-Ryan, 2014), channels attention to lack of reading as the primary cause of negative market outcomes, and focuses on an "average American" consumer while neglecting those with different linguistic or cognitive needs. It also galvanizes doctrines presupposing strong reading norms, such as the duty to read (Perillo, 2009), strict textualism (La. Code Civ. Ann. art. 2046, 2011), and shrinkwrap theory (Hill v. Gateway 2000, Inc., 1997). In an era of personalized "smart readers" capable of tailoring texts to individual sociocultural, linguistic, and cognitive needs, regulatory efforts should shift from pursuing a one-size-fits-all plain language ideal to addressing specific vulnerabilities and assisting at-risk individuals (Arbel & Becher, 2022; Kolt, 2022). This nuanced approach recognizes consumer diversity and the evolving technological landscape, potentially offering more effective protection than broad plain language mandates.

There are also broader lessons about the law's reliance on folk empiricism and the potential risks of well-intentioned but misguided reforms, such as the ALL-CAPS myth (Arbel & Toler, 2020). They highlight the need for evidence-based policymaking in consumer protection, cautioning against the allure of seemingly easy fixes that may divert attention and resources from more effective interventions. There are also a number of methodological contributions, concerning the use of machine learning tools to read and classify contracts, and of course, the development of a large trove of contracts for the benefit of future research.

The remainder of this Article proceeds as follows. Part I explores the Plain Language Movement and its four central tenets, critically examining the assumptions underlying the push for simpler contract language. Part II describes the novel dataset compiled for this research, detailing the diverse array of contracts analyzed and the innovative machine learning methods used to classify them. Parts III and IV present the core empirical findings, evaluating the readability of contracts relative to other genres and assessing how contract readability has evolved over time. Finally, Part V concludes with a discussion of the implications of this study for consumer protection law, market regulation, and the future of contract design in an era of technological change.

By challenging fundamental assumptions about the readability of legal agreements and consumer literacy, this Article aims to stimulate a reconsideration of how we approach the language of the law and its interaction with those it governs. It calls for a shift away from simplistic notions of plain language towards more nuanced, technology-aware strategies that can effectively address the real challenges consumers face in modern markets.

I. THE FOUR TENETS OF PLAIN LANGUAGE

The study of the plain language movement is a study of paradoxes (Schriver, 2017; Felsenfeld, 1981-1982). This movement is at the same time politically invisible and politically powerful; it is a populist movement that is endorsed by many elites; it is a consumer movement, but it is at least rhetorically endorsed by many businesses (Hoffman, 2018; Masson & Waldron, 1994); it is on the agenda of both Republicans and Democrats; it is a global phenomenon but it originates from a specific American post-War moment (Masson & Waldron, 1994; Zödi, 2019; GDPR, art. 4(1); Miyazaki, 2004); and it reached the height of its effectiveness long after its scholarly influence waned. The movement found itself in the 1990s in a series of heated debates, and then became part of the general and mostly undisputed scholarly commonsense in the aughts and continuing to this day with no loss of vigor. Despite of its paradoxes, and maybe even because of them.

A few facts explain some of the movement paradoxes. The movement emerges in the 1970s (Stabler, 2013; Flesch, 1981), finding leadership in plain language advocates such as Rudolph Flesch who also worked for the FTC (Flesch, 1981). This is an important moment for consumer protection. This is exactly the time that FTC substantive regulatory powers decline, as a result of the 'kid vid' debacle (O'Reilly, 2018; Herrine, 2021). Perhaps this instilled a sense in reformers that the most winnable battles lie in process rather than substance, in 'science' and pragmatics rather than values and politics. It is also a period of growing concerns about literacy and educational inequality, with the 1983 "Nation at Risk Report" sending shockwaves (Guthrie & Springer, 2004). A harbinger, no doubt, this report was false (Guthrie & Springer, 2004); but its effect marked a long-lasting "moment of angst" about American schools (Kamenetz, 2018). This instilled in the movement the sense that Americans are deeply uneducated. Lastly, it is the era of Ralph Nader, and his influential "battle against smoke-filled rooms populated only by well-heeled insiders" (Wagner, 2010). Nader's advocacy linked consumerism and freedom of information (Pozen, 2018), making it a powerful rallying cry for reform.

From all of that emerges a movement that is proudly technocratic, abjures any overt political manifestations, and is mostly dry of theory or jurisprudence. Much of its wisdom is reduced to manuals, guides, and grammar books. In them, one can find a certain folksy grace, with all the little jokes about obfuscating lawyers, clumsy firms, and elitist regulators. What is difficult to find there is much theory: these books are instead full of lengthy reprimands of those who use 'in order to' when they could have used 'to' (Wydick, 2005). It must be admitted, however, that the movement doth protest too much. There are clear jurisprudential and political elements to it. In particular, there are echoes of legal realism in the movement's irreverent disregard to high form and legalistic traditions. One continuously hears echoes of Judge Cardozo saying, "[t]he law has outgrown its primitive stage of formalism when the precise word was the sovereign talisman, and every slip was fatal." And rather than a-political, the movement is continuously endorsed by politicians of all stripes, to serve

transparency, populist, and sometime obstructionist agendas. Former Vice President Al Gore would hold monthly "No Gobbledygook Awards" and even argued that "plain language is a civil right." President Obama signed the Plain Writing Act of 2010, building on a pervious memo by President Clinton. Republicans too like it. Congress recently passed a law initiated by republican senators demanding that every agency rule to include a short plain language summary. Florida Governor Ron DeSantis, in his quarrels with social media platforms, signed a bill requiring that they use plain language in their consumer decisions, and Obama's law was so popular among Republicans that broke ranks within tea party, passing the house 386-33.
This Part takes this movement seriously. It identifies, explains, and ultimately critiques its four essential tenets. These tenets are motivated by the movement's technical leanings, but upon close examinations they all rely on unexamined assumptions that become doubtful upon close analysis.

1. People Do not Read Contracts *because* of Legalese

The first tenet of the movement is this: people do not read as much as they should *because* of the difficulty of language. As one sympathetic author described the literature, "That [p]lain English is something to be desired in legal writing . . . is something taken almost as an article of faith in legal writing circles" (Gallacher, 2013).

Both parts of this proposition are doubtful. There is reason to think that the no-reading problem itself is at least somewhat exaggerated (Gillette, 2004; Kornhauser, 1976; Elvy, 2018). When looking at actual practice, people do read some contracts some of the time (Schmitz, 2010; Stark & Choplin, 2009; Hillman, n.d.; Becher & Unger-Aviram, 2010; Plaut & Bartlett, 2012; Listokin, 2010; Hoffman, 2018; Davis, 2013). And even when people do not read "the" contract, this means much less than advocates charge. Firms advertise warranties, buyers inquire about return policies, and reputation markets are well integrated into the shopping process (Ayres & Schwartz, 2013). While there is compelling evidence that some online contracts are not read (Bakos et al., 2014), it is even doubtful that reading consistently helps consumers. Based on her empirical findings, Tess Wilkinson-Ryan started wondering whether it might serve consumers to read *less* (Wilkinson-Ryan, 2017; Zamir, 1997).

Second is the proposition that legalese is a major obstacle to reading. As noted, there are many sensible reasons why ordinary people quote for their reading aversion, such as the disutility of reading a document one cannot change. But in legal scholarship, there are strong condemnations of legal language. The source of complaints about legal language are old (Swift, 1726/2021; More, 1516/2021; Gopen, 1987), but the plain language movement is usually dated to the late 1970s (Wydick, 2005). Its most immediate roots are usually traced to the 1963 book, *The Language of the Law*, which served as a condemnation of the use of legalese and its remoteness from everyday speech. From then, the movement endorsed the idea that people do not read contracts because of legalese. This led scholars, beyond plain language reforms, to advocate for the

policing of the substantive terms of the transaction itself (Bar-Gill, 2004; Korobkin, 2003; Zamir & Farkash, 2015), non-enforcement of boilerplate terms (Rakoff, 1983; Radin, 2014), requirements of extended and conspicuous disclosure (Hillman, 2006; Hillman & Barakat, 2009), nonenforcement of difficult contracts (Rustad, 2022), alongside more general skepticism of consumer consent and its merits (Oman, 2016; Oman, forthcoming 2023).

2. Simplifying Language Would Help Consumers Read

The second tenet of this movement is that making language simple would matter. Plain-languagers believe that by turning legalese to plain language they would increase consumer wellbeing improving textual information retrieval, comprehension, and usability (Federal Plain Language Guidelines, 2011). In line with their technocratic bent, the source of unreadability is often not cast in terms of strategic obfuscation or the professional self-interest of lawyers (Bentham, 1843). Rather, legalese is simply a matter of bad craftmanship (Asprey, 1996). It results from inartful copyediting by "mediocre writers" who are accultured to write bad prose (Garner, 2001), and if lawyers want to earn the respect of the public (Asprey, 1996; Hathaway, 1994), they should pay sufficient care and attention to their products (Wydick, 2005).

In making these claims, plain languagers brush against three conceptual critiques: lossy compression, the length-simplicity trade-off, and reader preferences (Aiken, 1960). The first objection, in other words, is that legalese is valuable. The translation and summarization of legal into plain language is a 'lossy' process, in its information theoretical sense, meaning that some information is necessarily lost along the way (Fitriya et al., 2017). The word 'chicken' carries different meanings in different contexts, and precision writing will often require fine, detailed distinctions between broilers, stewers, fryers, and then further between game chicken and Cornish game. Using plain language, as Crump and others have argued, can be detrimental to the preservation of a complete and accurate record (Crump, 2002; Stark, 1994). Advocates deny that simplicity comes at the expense of precision and argue instead that plain language actually invites precision (Kimble, 1995).

Advocates are probably right that precision arguments are overstated. There is often a way to restate sentences in plain language, adding simple prose to explain every technical term. But this leads to the second concern. The additional explication comes at the necessary expense of length. To explain "FOB origin" in words accessible to an eighth grader—and much more so to a sixth grader, as some suggest (Rustad, 2022)—would require a good deal of exposition. However, bloated forms are themselves a readability hazard. Moreover, plain-languagers have always fended off against the allegation that they are using 'baby talk' by arguing that plain language is not necessarily juvenile (Kimble, 1995). Such a defense is harder to mount, however, when the goal is to make text accessible to juveniles. Dutch insurance providers were required to present more readable policies; the result of their concentrated efforts were

policies that were 30% easier to read but were 32% longer (Van Boom et al., 2016). Daniel Schwarcz attributes part of the length of modern insurance policies as a response to courts finding ambiguities in the older forms (Schwarcz, 2021). This means that often solving one readability problem—linguistic simplicity—will come at the expense of exacerbating other readability problem—e.g., text length.

The last conceptual critique is perhaps the most important one for the quantitative attempts to measure readability. The empirical evidence, explored later in this section, raises serious questions about the connection between plain language and its supposed effects. As one commentator summarized the literature: "the effect of language simplifications on comprehensibility is not a simple linear interdependence" (Zödi, 2019).

3. Readability Can be Quantified

The third tenet is that we can scientifically and objectively quantify the level of readability of texts. This belief led to a cottage industry devoted to such measurements of text readability (Zamanian & Heydari, 2012). Mostly developed by psychologists—rather than linguists—a few quantitative measures of text readability have risen to wide adoption, most famously the Flesch-Kincaid measure. For instance, Fitzsimmons et al. (2010) argue that "SMOG should be the preferred measure of readability when evaluating consumer-oriented healthcare material" (p. 295), calling SMOG the "gold standard" (p. 294).

The conceit of these measures is that formal aspects of the text are reliable indicators of its "readability." The construct of readability, in turn, is operationalized via formal textual features, ignoring the meaning strata of the text, and assign them a readability score (Crossley et al., 2019). The industry around readability measurement has grown to include service providers like "Readable," which offers team packages of software that evaluates text readability for $828 a year, and "Lexile," which sells a proprietary measure to governments and educators for $18 a year for individual educators. Since there are many such readability measures, the table below summarizes the most common ones and their logic.

Table 1 Summary of Readability Tests

TEST	WORDS	SENTE-NCES	SYLLABLES	COMPLEX / HARD WORDS	CHARACTERS
FLESCH-KINCAID READING EASE	✓	✓	✓		
GUNNING FOG	✓	✓		✓	
SMOG		✓	✓	✓	
COLEMAN-LIAU	✓	✓			✓
AUTOMATED READABILITY INDEX	✓	✓			✓
DALE-CHALL	✓	✓		✓	

Many plain language writers hedge: they recognize explicitly that these tests are insufficient and reductive, and that they do not always capture plain language fully (Asprey, 1996a). However, in other contexts, they are less careful, claiming that they have "hard evidence" of effectiveness (Asprey, 1996b). These are the exact tools they use to measure plain language and declare a readability gap. These tools are also widely used in practice by policymakers, educators, government agencies, commercial entities, and researchers (Plavén-Sigray et al., 2017).

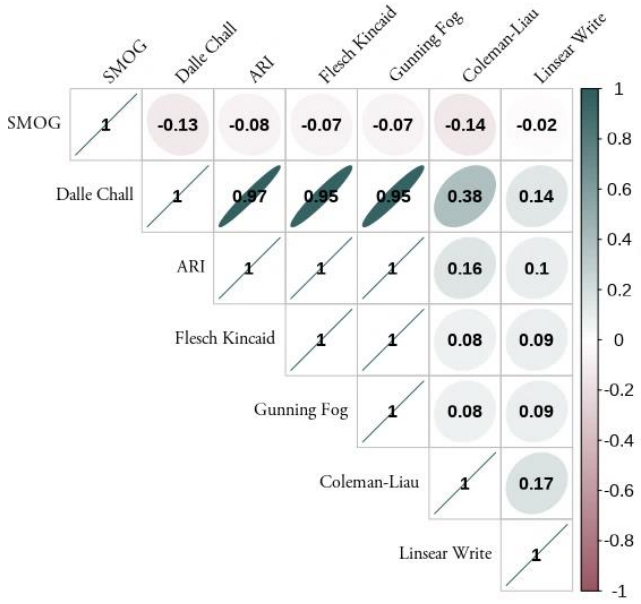
To assess the validity of these measures, it is crucial to examine both their internal validity and external validity. In other words, how well do these tests measure what they purport to measure, and how well does this construct correlate with real-life outcomes? The reality is far more complex than often assumed.

Readability tests are *not* internally valid. As noted, unlike linguists who focus on lexical sophistication, syntactic complexity, and discourse structures (Crossley et al., 2019), the plain language tests focus on formal aspects of the text which they believe to be related to the same construct measured by linguists.

This assumption leads to a testable prediction. Because “"the readability formulas all purport to measure similar constructs" (Crossley et al., 2022, p. 91)” we should expect that their results to be internally correlated.

To test this assumption, I calculated the readability scores of the entire Gutenberg project (consisting of 50,000 books) (Parrish, n.d.), and then evaluated the inter-correlation of these tests, reported in the figure below.

Figure 1 Correlation Matrix of Readability Tests on the Guttenberg Library



The analysis of readability tests reveals significant inconsistencies and reliability issues. The correlation matrix of various tests on the Gutenberg library demonstrates that Flesch-Kincaid, Gunning Fox, and Dale-Chall have minimal correlation with Coleman-Liau, Difficult Words, Linsear Write, and SMOG tests. This lack of correlation suggests that these tests may not be measuring the same construct, or that 'readability' itself may not be a unified concept (Crossley et al., 2019). While this doesn't necessarily invalidate all tests, it fails to indicate which ones are truly valid. Conversely, the near-perfect agreement among Flesch-Kincaid, Gunning Fox, and Dale-Chall implies that they may be redundant, offering little independent information.

A more concerning issue emerged during testing: these measures lack consistency even within *themselves*. In an experiment, I collected 10,000 paragraphs from contracts and ran the same readability tests on them, using various python libraries. [1] Even though these libraries ostensibly use the same mathematical formulae, Flesch-Kincaid scores varied by an average of 4.6 grade

[1] The average gap was calculated as $\mathbb{E}(S_{max} - S_{\min})$, where S is the score for a given test. The average gap for Flesch-Kincaid was 4.6 (median 3.77. If max grade is capped at 20 years, the average becomes 3.36).

levels depending on the specific library used. This variability suggests that researchers or drafting firms could potentially manipulate outcomes simply by choosing different implementations of the same test.

The root of these inconsistencies lies in the technical implementation of the tests. As outlined in Table 1, these measures rely on counting sentences, words, and syllables. However, contrary to common assumption, there's no universally agreed-upon method for dividing text into these units. Sentences, for example, are not merely collections of words ending with terminal punctuation or line breaks. This ambiguity can lead to absurd results, such as calculating that the Star-Spangled Banner requires 46 years of schooling to read, due to its unique verse structure. Similar issues arise in contracts with headings, clause titles, bullet points, definitions, and recitals. File conversions from their web or scanned format usually result in many sentences that are broken in the middle.

Word definition also presents challenges. While it might seem straightforward, complications arise with compound words (e.g., "New-York," "mother-in-law"), contractions (e.g., the Southern gem "y'all'd've'd"), and stylistic choices (e.g., "W I T T E N E S S E T H"). Syllable counting proves even more problematic, varying based on regional pronunciations and linguistic oddities. For instance, the syllable count in "Caramel Mayonnaise with Chocolate? Interesting!" strongly differs depending on the speaker's local accent. The English language's peculiarities, such as dropped syllables in words like "comfortable" and "temperature," further complicate matters. Digital conventions add another layer of complexity, raising questions about how to treat elements like "www.google.com" in syllable counts.

Previous research has largely overlooked these issues, necessitating a cautious approach when interpreting earlier findings. The choice of a specific test can significantly influence a study's results. While the present analysis incorporates measures to mitigate these issues, it's clear that no perfect solution exists. Readability tests, despite their widespread use, remain flawed and unreliable metrics.

Even if one is willing to overlook issues of internal validity, the question of external validity—how well each readability test performs in measuring actual text readability—presents a complex picture. Despite Kincaid's claim of accuracy "within a single grade level," evidence is mixed at best (Kincaid et al., 1975). Some studies report strong correlations, while others find none (Zamanian & Heydari, 2012). For instance, a study of 118 foreign English speakers found no statistically significant correlation between readability tests and comprehension, with better readability scores paradoxically leading to worse outcomes (Greenfield, 2004). In scientific research, less readable abstracts are sometimes more frequently cited (Plavén-Sigray et al., 2017).

This raises the counterintuitive possibility that 'readable' texts may not always be preferred by readers. A study on simplified medical disclosures revealed that overly simple language was not well-received (Klare et al., 1955). On reflection, it would seem that readers appreciate texts matching their own level,

rather than the simplest possible version. This might explain the curious finding that, in scientific writing, *less* readable abstracts are more frequently cited.[2]

Recent work by Crossley and colleagues provides more nuanced insights. They found modest correlations (0.3-0.4) between subjective comprehension, reading speed, and scores from Flesch-Kincaid and New Dale-Chall tests (Crossley et al., 2017). While these tests performed well in distinguishing between standard and simplified Wikipedia articles, this finding may be circular if Wikipedia editors use these tests. When measuring actual comprehension and reading speed of Wikipedia articles among undergraduate students, Flesch-Kincaid and New Dale-Chall showed weak explanatory power, accounting for only 6-14% of the total variance (Crossley et al., 2019).

Studies on *contract* simplification similarly yield mixed results. While Rudolph Flesch argued his formula worked well for contracts, subsequent research has been less conclusive (Flesch, 1979). Some studies show positive effects of simplification on contract readability, but often with unrepresentative samples or problematic design choices (Davis, 1977). One study of older participants (N=25) found that even substantial simplification efforts did not affect comprehension or subjective ratings of text helpfulness (Campbell, 1999). Research on trust and perceived transaction fairness found no effect from improved readability, though it did increase the likelihood of post-sale clarification requests (Bloon, 2016). A study with 96 participants found no relation between readability formulas and contract comprehensibility, with the least readable forms sometimes yielding the highest correct answers (Schmitz & Pavillon, 2020). The authors concluded that: "[i]t is therefore likely that our assumption that readability largely equates to comprehensibility . . . is wrong." (199). Finally, a study found that of many interventions ion contract form, the only one that increased comprehension was making them shorter. (Stark & Choplin, 2009).

In what is probably the most comprehensive studies to date, involving 1,542 participants, researchers exposed participants to 'easy' (10th-grade level) and 'complex' (college level) contract versions. While participants perceived the easy version as more understandable, this did not affect their behavioral responses to contract breach or likelihood of seeking advice, filing complaints, or taking legal action (Van Boom, 2016).

In conclusion, the evidence suggests that while quantification of readability correlates with some measures of comprehension, this relationship is weak, inconsistent, and unreliable. In the context of contracts, there is insufficient evidence to conclude that improved readability scores significantly enhance consumer understanding or decision-making. Even under the most optimistic interpretations, large effects are unlikely. These findings highlight the need for caution when relying on readability tests and emphasize the complexity

[2] Lennart Ante, *The Relationship Between Readability and Scientific Impact: Evidence from Emerging Technology Discourses*, 16 J. INFORMETRICS, 16 (2022) (using Flesch-Kincaid, Smog, and ARI).

of measuring and improving text comprehension, particularly in legal and consumer contexts.

4. Americans are Functionally Illiterate and Depend on Plain Language

The final tenet, and in practical terms the most important one, is that Americans require simplification because their literacy levels are very low. Readability literature has emphatically stated that half of American adults cannot read at a level beyond that of an eighth grader (Barnes, 2017; Dumas, 2000; Green, 2013; Plate, 2010; Rustad, 2022; Coughlin, 2015; Kuczmarski et al., 2016; McCarty & Rogers, 2012; Benoliel & Becher, 2021). One study advocates for a "ninth grade level" (Smith & Lane, 2018), and a number of studies even advocate the use of a "sixth grade" level (Weaver, 2023). This pervasive idea made its way into government reports, where it is stated as an established fact, and even into Wikipedia (Wikipedia, "Literacy in the United States", n.d.; Wikipedia, "Plain Language", n.d.). An ironic but illustrative anecdote, the fact checking website Snopes, commissioned a check whether "more than half of Americans read below 6th-grade level"—and find reports this as true (Dapcevich, 2022).

This is a big claim with profound implications. If Americans can read no better than 11 or 13 year olds, this means that we should expect our peers, friends, and neighbors to struggle with any written material not fit for young teenagers. It should raise concerns about America's ability to produce and innovate, and perhaps even about the viability of our democracy. Fortunately, this enduring myth is false.

The following table summarizes the latest US Census Data on educational attainment of American adults (US Census, 2022).

Table 2 Educational Attainment of Adults 25 and Older. U.S. Census Data. 2022

Racial Group	College or Higher	High School Diploma	9-12th Grade	5-8th Grade	Remainder
White	63%	25.8%	5%	2.6%	3.5%
Non-Hispanic White	67.9%	24.9%	3.6%	0.9%	2.7%
Black	56.4%	31.2%	7.6%	1.6%	3.2%
Asian	73.5%	17.8%	3.3%	2.9%	2.6%
Hispanic (of any race)	42.5%	30%	10.9%	9.6%	7%
All	62.8%	25.9%	5.2%	3.1%	3%

As the table shows, the average American adult has over thirteen years of schooling, with 77.5% having at least a high school diploma. Only 0.1% of

Americans fall into the 5-8 grade level category. The Program for the International Assessment of Adult Competencies (PIAAC) finds that 79% of Americans have mid or high English proficiency, with only about 20% at low literacy levels (Level 1), and an even smaller percentage defined as illiterate (below Level 1).

Years of schooling may not accurately reflect students' actual learning outcomes. The Program for the International Assessment of Adult Competencies (PIAAC), a leading comparative literacy measure, directly assesses individuals' practical literacy skills. This test reveals that 79% of Americans demonstrate mid to high English proficiency (National Center for Education Statistics [NCES], 2019). This finding is consistent across various racial and ethnic groups (NCES, 2019). Approximately 20% of Americans fall into the Level 1 category, indicating low literacy levels (NCES, 2019). However, only those performing below Level 1 are formally classified as illiterate (NCES, 2019). There is no evidence, in either formal education or actual achievement, that American adults have arrested development at the 8th grade level.

Three main sources are often cited to support claims of low literacy. The 2003 National Assessment of Adult Literacy (NAAL) survey is one. On close inspection, however, it not only fails to support the claim but strongly discourages the use of grade levels to measure literacy, calling them "arbitrary standards" (NCES, 2002; White & Mansfield, 2002). What it does find is that the majority of adults (66%) possess an intermediate or higher level of literacy, meaning that they can at the very least "perform moderately challenging literacy activities" (NCES, n.d.). They are capable of reading "information in dense, complex documents and mak[ing] simple inferences about the information" (NCES, 2006). Over 93% of people in this group rated their ability to read in English as "well" or "very well" (NCES, n.d.), a far cry from what we would expect from those who struggle daily with literacy. Even individuals reading below these levels may not necessarily struggle with literacy: notably, 13% of graduate and doctoral students do not meet these standards (NCES, n.d.). It's worth noting that the GED is intentionally designed such that 25-30% of test takers will fail it (Baldwin et al., 1995).

A different source for such claims comes from a Gallup report (Rothwell, 2020), which is based on the PIAAC test previously described. This report indeed concludes that 50% of American adults are "at least partially illiterate" (Rothwell, 2020, p. 6). However, this alarming conclusion rests on a simple maneuver: the author of the report stipulates that "Adults who score below Level 3 for literacy are not considered proficient and are defined as at least partially illiterate in this study" (Rothwell, 2020, p. 6). This definition does not appear in the OECD's report (OECD, 2021). Rather, the Institute of Education Sciences describes level 2 as

> "[N]earing proficiency but still struggling to perform tasks with text-based information. *Such adults may be able to read print and digital texts, relate multiple pieces of information within or across a couple documents, compare and contrast, and draw*

> *simple inferences*. However, more complex inferencing and evaluation may be too difficult." (NAT'L CTR. FOR EDUC. STAT.)

Adults who are able to compare and contrast multiple texts can hardly be defined as partially illiterate, much less held to be at some specific grade level.

The last source comes from various studies in healthcare literature. This is quite a different context, where special attention must be paid to distress and physical disability. The studies in this literature tend to depend on relatively small samples, involve the comprehension of medical materials, and often use specialized readability measures (Logue et al., 2022). One central source in this field is an out-of-print textbook that offered a "rough approximation" that people scoring between Level 2 and 3 are at an "8th- to 9th-grade level" (Doak et al., 1996, pp. 2-3). This assertion is not sourced, and no modern work in education could be found to support it. Other studies cited in non-academic sources, such as Wikipedia, offer little pertinent support. For instance, one such citation is a one-page article in a magazine that reports a correlation between Flesch level below 8th grade and "readership scores" through an unidentified methodology (Feld, 1948).

The 8th grade myth conveniently ignores a wealth of contradictory evidence. In the international literacy context, the U.S. performs slightly above the OECD average, comfortably between England, Germany, and Canada (OECD, n.d.). If the U.S. were majority illiterate, it would imply that literacy is far more common globally than anyone believes. Statistics on book reading, news consumption, and social media use (which is text-dense compared to TV) suggest that most adults frequently interact with non-trivial English texts. In fact, 83% of Americans read at least one book over the last year, and 41% read more than six (Pew Research Center, n.d.). Moreover, the U.S. has improved on almost every measure of literacy over the last decades, with adults becoming more capable than their 1960s or 1970s counterparts.

Perhaps most tellingly, a study led by Stanford sociologist Sean Reardon reports that 75% of eighth graders read at or above the basic level for their grade, and 76% of 12th graders read at or above the basic level for theirs (Reardon et al., n.d.). The notion that half of America never progresses beyond an 8th-grade reading level is false—even for those currently in the eighth grade.

A simple sanity check would illustrate the point. If the U.S. were truly majority "partially illiterate," then given its international comparisons, this would paint a global picture of rampant illiteracy in the developed world. In fact, however, statistics on book reading, news consumption, and use of social media (which, at least relative to TV, is text-dense) suggest that most adults interact with non-trivial English texts with a very high frequency (Gelles-Watnick & Perrin, 2021). 83% of Americans read at least one book over the last year, and 41% more than six (Jones, 2022; see also Gelles-Watnick & Perrin, 2021; Watson, 2023). Moreover, the U.S. has improved on almost every measure of literacy over the last decades, with adults becoming more capable readers than their counterparts in the 1960s or 1970s (Shakeel & Peterson, 2022).

With this conclusion in mind, another thing should be clearly stated: it is legitimate and, in my view, quite desirable for policymakers to focus on non-average Americans, people with real vulnerabilities and literacy difficulties. It is also important to recognize that even adults with high literacy scores may find legal texts challenging. The plain language movement, however, has largely set its goals on the average American and their average grade level reading skills (White & Mansfield, 2002; Rustad & Koenig, 2013; Dyer et al., 2013; Ciocchetti, 2007).

This choice is all but natural. Literacy is a complex phenomenon and those who struggle with it are highly heterogeneous. Some may find reading challenging due to dyslexia, ADHD, or other forms of neurodivergence; for some English is not a first (or even fourth) language and others who speak one of its many dialects; some of them lack cultural capital while others have a visual deficiency; some were deprived of an education while others are aged, young or old. Moreover, the contexts in which people encounter contracts—whether stressful and hurried or relaxed and guided—can deeply influence their comprehension and decision-making process. To reduce readability to a single 'grade level' is to suppose there is a single metric that will be useful to all those who struggle.

But even if focusing on average Americans in this context is a sensible goal, the idea that the average American cannot read past the 6-8th grade level is false. It builds the wrong expectation, leads to wrong policy measures, and raises questions about the rigor and scientific merit of the movement as a whole.

II. DATA: AGGREGATION, PROCESSING, AND MACHINE LEARNING CLASSIFICATION

The Plain Language Movement's four tenets have been described and shown to be problematic. However, their central claim—that contracts and other legal documents are overly complex and generally unreadable—remains to be thoroughly examined. This study aims to rigorously test this assertion. Despite the identified issues with traditional readability tests, this research employs a modified, more robust version of these metrics. These enhanced measures address several of the previously discussed concerns, providing a stronger foundation for analysis. Using these improved tools, this study conducts the most comprehensive empirical evaluation of contract readability to date.

This section focuses on data aggregation, cleaning, and machine learning classification, offering rich detail as the large dataset presented here is made publicly available for future researchers. Additionally, this work contributes to the growing field of "legal NLP"—the processing of legal language by machines to address issues of access to justice, improve judicial processes, and enhance drafting (Arbel & Hoffman, forthcoming).

1. Datasets and Classification

1. Commercial Contracts

Publicly traded companies must report all material contracts, defined as "[e]very contract not made in the ordinary course of business which is material to the registrant" (17 C.F.R. § 229.601(b)(10)(i)). Materiality encompasses any issue that may significantly impact a company's financial performance and shareholder value, including cash or non-cash bonuses, severance arrangements, or change-in-control provisions. Filings may also be required for agreements containing specific financial arrangements like stock options, restricted stock units, or performance-based awards.

Material contracts, filed as "exhibit-10," typically accompany annual, quarterly, or important event filings (forms 10-K, 10-Q, or 8-K). To ensure comprehensiveness, every SEC filing was algorithmically reviewed for exhibit-10, yielding a dataset of 1.2 million contracts (Nyarko, 2021).

Data formatting is crucial, as improper parsing can distort readability analysis. To ensure consistency across various file formats (plain text, PDF, HTML), all documents were converted to plain text using Python libraries such as Beautifulsoup and PDFMiner (Ramachandran et al., 2022). This process resulted in well-formatted data, mitigating concerns about artifacts from faulty formatting.

The broad nature of "commercial contracts" necessitated a classification system to differentiate between various contract types. However, the volume of data made manual classification impractical. Previous research attempted to address this using hard-coded machine rules (Nyarko, 2019), which worked

reasonably well for many contracts but fell short with idiosyncratic agreements. Lawyers often employ considerable variation in describing their agreements, using casual language, industry terms, and occasionally introducing typos. I created 275 such rules myself, aiming to catch the most common types, but could only classify a quarter of the contracts.

To address this classification challenge, machine learning offers a promising solution. However, training a robust model typically requires a large dataset of labeled documents - a resource-intensive task when dealing with millions of contracts.

The approach chosen here leverages an important feature of legal documents: their built-in labels. The title of a contract, crafted by the drafting lawyers, serves as a reasonable indicator of its type (Uniform Commercial Code, 2001). This characteristic allowed us to sidestep the need for manual annotation, which would have been prohibitively time-consuming and potentially inconsistent.[3] However, relying solely on titles for classification would be shortsighted. Our goal was to create a model capable of discerning contract types based on their overall structure and content. To this end, a three-step process was employed:

1. Using a number of titles to create our initial labeled dataset.
2. Grouping similar titles into 10 leading categories (e.g., lending agreements and secured lending contracts were both grouped as "Credit, Debt, and Security Agreements").
3. Scrubbing these titles from the documents before training the model.

This approach ensures that our model learns to classify contracts based on their substantive content rather than superficial markers. The model must learn to identify the document type without access to the title. This method yielded a labeled dataset of 275,480 contracts across ten categories, providing a solid foundation for our machine learning model without the need for extensive manual annotation.

Table 2: Types of Pre-Classified Contracts

Agreement Type	Number
Credit, debt, and Security Agreements	58,586
Status Agreement	43,992
Insurance, Indemnity, and Coverage	11,448
Mergers, Alliance, and Investment Agreements	10,046
Ownership, Trust, and Governance	7,611
Property, Rights, and IP	25,515
Purchase or Sale Agreement	37,818
Services & Supply	23,248

[3] It's worth noting that while this approach is generally effective, it may not be universally applicable. In some areas of law, such as secured transactions, parties may attempt to mischaracterize the nature of their agreement. However, such concerns are marginal in the context of this study.

Settlement, Waiver, and Termination or Severance	12,689
Shares, Stocks, Incentives, and Options	44,527
Total	**275,480**

After preprocessing the dataset, a machine-learning model was trained for contract classification. A crucial step in this process is vectorization or embedding, which converts words into numerical representations that computers can process (Arbel & Hoffman, forthcoming).

Various vectorization methods were tested, including state-of-the-art transformer-based embeddings (Bert, DistilBert, and Legal Bert) and the more traditional TF-IDF vectorizer. Surprisingly, despite its simplicity and age, the TF-IDF vectorizer outperformed the more complex methods.

The TF-IDF method assigns numerical values to words based on their rarity in the corpus, emphasizing specialized words that are more indicative of the contract type. Common words like "and" and "today" are considered less important than specialized terms like "mortgage" or "at-will." The result is a per-document vector of numerical values, each representing the importance of a particular word to that document within the context of the entire corpus (Ramos, 1999).

This approach is particularly effective for contract classification because it naturally highlights the distinctive vocabulary of different contract types. For instance, employment contracts might have high TF-IDF scores for terms like "salary," "benefits," or "termination," while credit agreements might score highly on "interest," "collateral," or "default."

In contrast, embeddings assign numbers to words based on their "meaning," with closer numbers signifying related concepts (Pilehvar & Camacho-Collados, 2021). Creating embeddings for the sample contracts took approximately 20 hours on an A100 GPU, accessed through the University of Alabama's High-Performance Computing Servers.

For the training phase, a standard 80-20 split was employed. Several classification models were trained, including fine-tuning an LLM model (LLaMa) (Touvron et al., 2023), deep neural networks, and various shallow classifiers. The Light Gradient Boosting Machine emerged as the best performer, considering both accuracy and inference time.

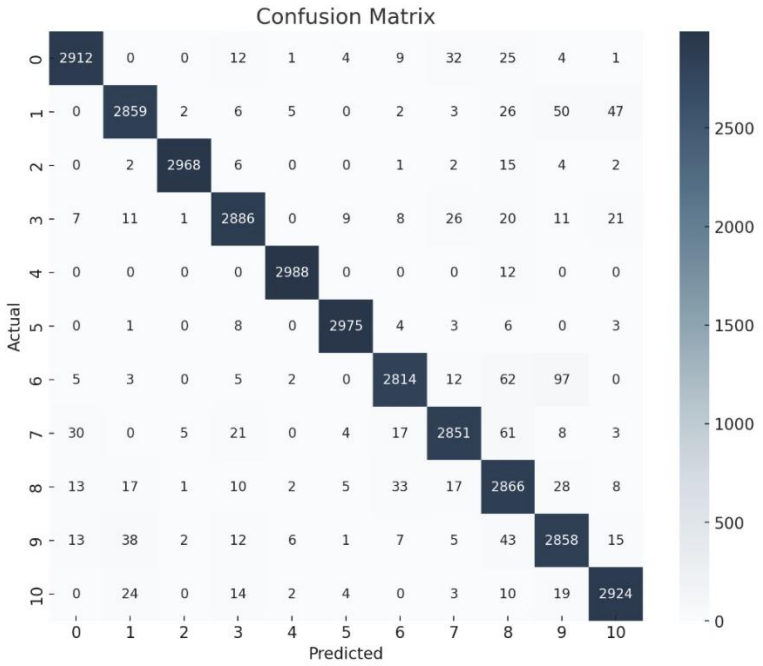
The Light Gradient Boosting Machine builds an ensemble of decision trees, each a series of if-then rules that split contracts based on their TF-IDF vector features. This approach allows the model to capture complex, non-linear relationships in the data, effectively learning which combinations of words and their importances are most indicative of different contract types. For instance, it might learn that a high TF-IDF score for "employment" combined with a moderate score for "stock options" strongly suggests an executive employment agreement.

The model's ensemble nature also provides robustness against overfitting, crucial when dealing with the nuances of legal language. Its lightweight design ensures computational efficiency, enabling faster training and

inference times compared to other boosting methods. These characteristics make it particularly well-suited for our large-scale contract classification task.

The model achieved a balanced accuracy score of 96.67% and F1 scores ranging between 93% and 100% for some contract categories, indicating high reliability.[4] These scores represent accuracy on the types of contracts included in the training and test datasets, and is described in the confusion matrix below:

Figure 2 Confusion Matrix



The "Confusion Matrix" can be interpreted as follows: the shaded box indicates the number of documents correctly classified, while each column explains how the remaining contracts were misclassified. For instance, we observe that the vast majority of contracts were correctly classified. Notably, credit and debt agreements were most frequently mistaken for Sale and Purchase agreements, which is logically consistent given that various credit arrangements often underlie sale agreements.

It's crucial to note, however, that these accuracy scores are representative only of the contract types included in the training and test datasets. To evaluate the model's performance on a broader range of contracts, a blind audit of 100 contracts was conducted. This audit yielded a more modest but still respectable balanced accuracy of 78%.[5]

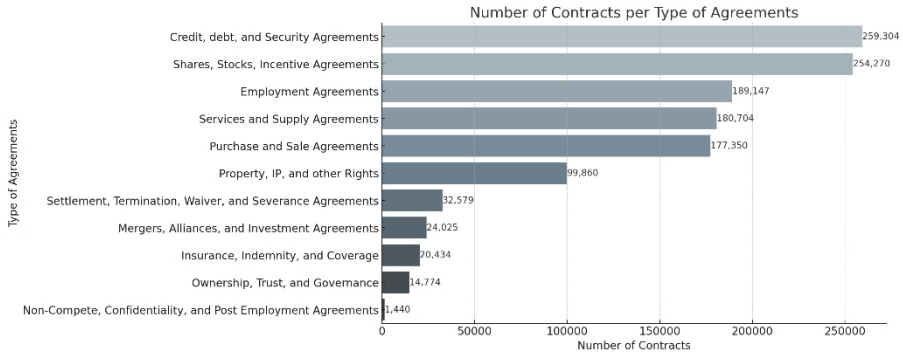
[4] On the interpretation of F_1 scores, *see* Nyarko, *supra* note **Error! Bookmark not defined.**, at 32. Using a deep learning model, it was possible to achieve an F_1 score of 96%, but that came with heavy inferential costs. All code and data will be made available upon publication.

[5] I am thankful to Julian Nyarko for the suggestion.

This additional step provides a more comprehensive assessment of the model's generalizability and robustness across diverse contract types, offering a more realistic picture of its performance in real-world applications

Applying the model to the total collection, the following figure describes the distribution of contracts in each category:

Figure 3 Number of Commercial Contracts per Category



2. Credit Contracts

Section 204 of the Credit Card Accountability Responsibility and Disclosure Act of 2009 (CARD Act) mandates that all consumer credit providers must submit their card agreements to the Consumer Financial Protection Bureau (CFPB) (Credit Card Accountability Responsibility and Disclosure Act, 2009). As a result, the CFPB has maintained a database of credit card agreements since the third quarter of 2011 (Consumer Financial Protection Bureau, n.d.).

This dataset is of immense importance for studying consumer contracts, as it captures nearly the entire landscape of credit card agreements in the United States. Considering that 80% of Americans use credit cards and each consumer possesses, on average, between three and four credit cards (Foster et al., 2021; Pino, 2022; Keys & Wang, 2016), the total number of credit cards in circulation is estimated at 485 million (Bureau of Consumer Financial Protection, 2021).

While the dataset is accessible, its organization presents challenges. No data was collected in 2015 (Bureau of Consumer Financial Protection, 2015), the file mapping agreements to issuers is incomplete, and there appears to be little quality control for submission formats. Some files are skewed scanned documents, while others are corrupt, duplicates, or empty.

Despite these challenges, most files were made usable through preprocessing. After downloading and removing duplicate files, a total of 95,071 agreements were obtained. The PDFMiner Python library was used for file conversion. Following several quality assurance rounds, PDF repair, and checks for faulty conversion or non-English language, the resulting dataset comprises 93,078 readable and processable credit agreements. It's worth noting that some of these documents are simply disclosure pages and do not contain the full terms

of the agreement. While this may affect readability assessments, these documents were included due to their importance to the transaction.

3. Privacy Policies

Data on privacy policies was sourced from Princeton Research (Amos et al., 2020). The dataset was originally obtained by identifying a set of target websites that appeared in the Alexa Top 100K list from 2009 to 2019. This range of websites was chosen to reach into the long tail of the web while avoiding overwhelming computational challenges. The timeframe was divided into six-month intervals, with snapshots of each website taken from each interval. Alexa's daily archives were utilized to identify the rankings closest to the midpoints of these intervals, which led to the identification of over half a million unique websites.

The researchers then queried the Wayback Machine's API to determine the list of homepage snapshots available for each website within these six-month intervals. The snapshot closest to the midpoint of the interval was selected when multiple snapshots were available. In cases where a website made it to the Top 100K list only in a particular year, snapshots from other years were also included, extending as far back as 1996.

For the purposes of the current research, their data was acquired and organized by year. Where policies for certain years were missing, the previous year's policy was used, under the assumption that the policy remained unchanged in the interim. Overall, the dataset analyzed contains a little over 630,000 privacy policies.

4. Franchise Agreements

Under federal and state law, franchisors are required to provide prospective franchisees with a number of disclosures, known as the Franchise Disclosure Document (FDD) (Benoliel & Zheng, 2018). These documents contain twenty-three items related to the identity of the parties, finances of the franchise, fees, and the franchise agreement itself (item 22) (Disclosure Requirements and Prohibitions Concerning Franchising, 2007).

States differ significantly in whether they require franchisors to also list their documents with the state. However, four states have a publicly open registry, and several others rely on the services of a third-party organization called NASAA (North American Securities Administrators Association, n.d.). For the purposes of this project, the different public databases were accessed and all available franchise disclosure agreements were scraped. The resulting dataset is significantly larger than previous studies in this area (Benoliel & Zheng, 2018). The following table summarizes the number of available documents by state:

Table 3 Franchise Disclosure Documents by State

Agreement Type	Number
Illinois	23
Rhode Island	46
Maryland	111
South Dakota	165
Minnesota	483
Virginia	749
North Dakota	1,370
New York	1,387
Indiana	1,571
Wisconsin	1,646
California	5,051
Total	12,602

***

In summary, after preprocessing, the final dataset comprises 1,935,680 contracts, of which 1.2 million are commercial contracts and 735,680 are consumer agreements (including franchise disclosure documents, which are considered hybrid agreements). This extensive corpus forms the foundation for our analysis of contract readability.

The representativeness of this sample merits careful consideration. Three key points should be emphasized:

(1) Most consumer agreements in this sample are standardized forms rather than individual contracts. A single form may represent tens of thousands of distinct contracts, as is the case with credit agreements. Consequently, the dataset effectively represents many millions of contracts covering significant aspects of consumer life.

(2) It is crucial to acknowledge the limitations of this sample. While extensive, it does not encompass the full spectrum of consumer agreements. The contracts in this dataset are predominantly subject to stringent regulatory oversight, which may influence their readability. For instance, credit card agreements, a significant portion of our sample, fall under the purview of the Truth in Lending Act (TILA), which mandates clear, conspicuous, and somewhat standardized disclosures. This regulatory context might contribute to the finding that these contracts present a level of difficulty comparable to that of reading daily news.

(3) Many other types of consumer contracts are not represented in this sample, such as agreements for online social media platforms, consumer-to-consumer transactions like used car sales, or retail purchase agreements between consumers and firms. Such contracts often operate under less rigorous federal clarity and standardization mandates, potentially resulting in different readability characteristics.

The limitations of this study should be evaluated against the empirical literature that was used to support the plain language reform agenda (Amos et

al., 2020). As noted earlier, previous studies relied on samples no larger than 500 contracts, from highly selected sources, representing tiny slices of the consumer contract space.

Given these considerations, the dataset should be interpreted as representative of a significant, but not exhaustive, portion of consumer contractual activity. Caveats notwithstanding, this analysis relies on a larger, more comprehensive understanding of the contract space than previously available.

2. Data Analysis

The discussion in the previous section highlighted the utility and limitations of readability tests. The strategy employed in this study involves strict preprocessing, averaging within tests, a new measure of readability (CRM), and benchmarking.

Rigorous Preprocessing. Preprocessing refers to a stage in data analysis where the text is cleaned to ensure uniformity. Here, the texts were cleaned of nonstandard characters, HTML tags, and other non-textual elements. Crucially, documents were broken down into sentences using two sentence tokenizers (Spacy and TextBlob) to ensure high accuracy in sentence tokenization The sentence tokenization was challenging, because when working with a large dataset of complex documents there many formatting options for sentence and clause structure that drafters use. Erroneous tokenization could make common clauses appear as if they require more than 90 years of schooling to read.[6]

Averaging within tests: As discussed, the readability tests are extremely sensitive to implementation, with a 4.6 average distance of the same test on same texts. To mitigate these issues I used a new approach: I collected a number of popular implementations of each tests and averaged their results. This helps control variability and may also increase accuracy, but most importantly, it constrains the researcher (or the regulated entity) in a choice of a test that would serve their desired narrative. (this is not to imply that previous research was deliberate in its test choice).

CRM, Averaging across Tests: Because each test (ideally) captures a different aspect of readability, and because there are so many tests, I created a summary measure called Composite Readability Measure, which is an average of each test average scores.[7] This novel measure is an average of averages across the different implementations of different readability scores.

Benchmarking. Because of the skepticism about the accuracy of these tests, I offer here a comparison of these texts vis-à-vis other texts adults read daily.

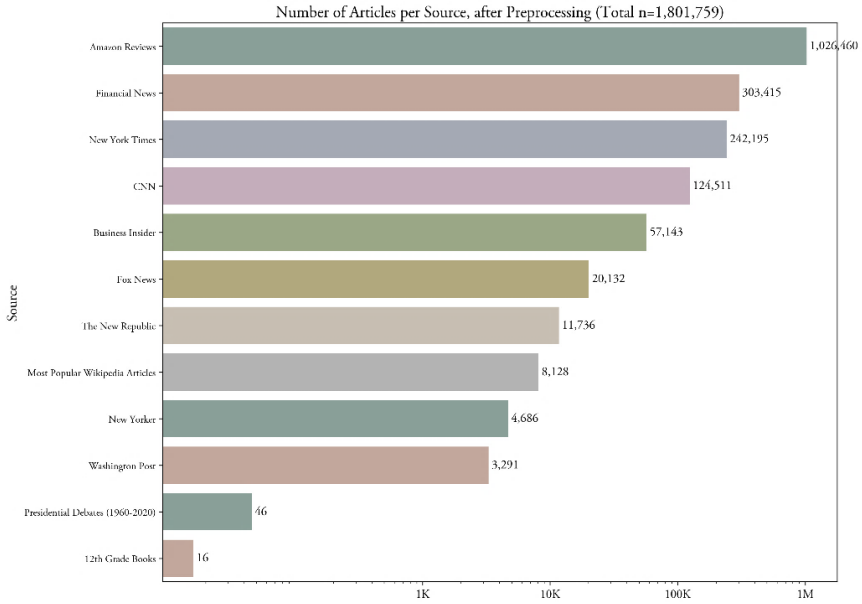
---

[6] See https://github.com/yonathanarbel/Contract_Dataset/blob/main/Paper_References/example_clause.txt

[7] Some files resulted in problematic scores, so I capped the maximum score at no more than 25 years of schooling, and restricted negative number of schooling years.

The idea is to not fixate on grade level in the abstract but to see differences between contracts and the texts that people read everyday. I created a dataset with 1.8 million texts by CNN, Fox News, Amazon product reviews, The New Yorker, a collection of stories on financial news, top-read articles on Wikipedia, transcripts of presidential debates from 1960-2022, and the text of popular books read by 12 graders.[8] The following figure summarizes this dataset:

Figure 4 Number of Articles per Source



These sources attract millions of daily readers from diverse demographics (Pew, 2020), with 69% of Americans reported common news consumption on websites (and 32% in print) (Pew, 2021). Adults spend, on average, 10 minutes a day readings newspapers and 15 minutes consuming news online (FCC, 2011). Critically, these are *voluntary* readers—people who choose to consume them out of interest, rather than as part of some transactional necessity, and may sometimes even *pay* for their consumption. On average, 75% of visitors to news websites spend there over 6 minutes. As such, these diverse sources provide a fuller sense of what we can broadly expect of most adults—again, bearing in mind that there are groups with special linguistic or other needs.[9]

[8] The main inclusion criteria was large corpora of texts that could be downloaded without payment and that reflected common texts read by American adults. The Wikipedia collection contains 10,000 random articles. Financial news were sourced from https://www.kaggle.com/datasets/jeet2016/us-financial-news-articles. It includes these publishers: Bloomberg.com, CNBC.com, reuters.com, wsj.com, fortune.com

[9] It is worth noting that those with atypical needs do not belong in some homogenous category of needs. While plain language is a remedy for some, it may do little for others. The problem is that addressing the needs of people in this group sometimes present conflicting demands on the drafter, regarding the choice of text versus audio, choice of language, font style and color, conflicting demands on text presentation, and possibly demands for audio-visual presentations.

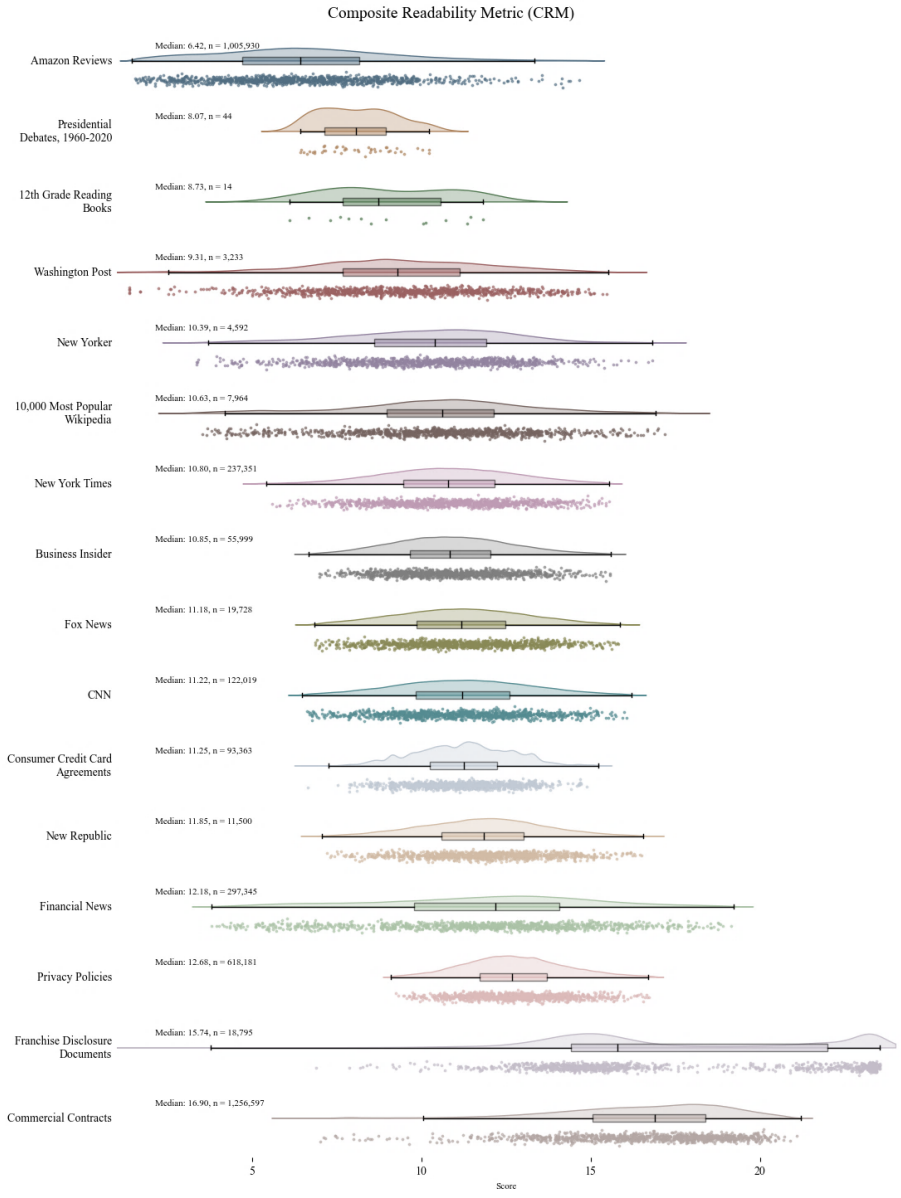
III. FINDINGS: CONTRACTS AND THE COMMON READING HABITS OF AMERICANS

This Part turns to describe the results of the readability analysis. It starts with comparing the readability level of contracts to texts we know American adults read on a daily basis, it then moves to examine how readability patterns have evolved over the last decades. The implications of these findings are discussed in the next Part.

1. The Readability of Contracts and Texts Americans Habitually Read

The following Figure reports the average results of all the various readability scores on the datasets. The first Figure uses the composite measure of all the readability tests (CRM), which as noted, averages across the leading readability metrics.

Figure 5 Composite Readability Metric (CRM) Scores by Type of Agreement



This "raincloud plot" describes both the median (dot in the middle) and the distribution of text difficulty, using the various readability tests. It is sorted by median difficulty. The simplest texts on the top are Amazon Reviews (median grade level 6.42) and transcripts of Presidential Debates (8.07). The most difficult and complex texts are the various commercial contracts (16.9) and the franchise disclosure agreements (15.74). In the middle are various common texts, such as CNN news (11.22) and Fox News (11.18). We can see that credit card agreements (11.25) are nestled in the middle, with privacy policies (12.68) are a tad harder, but not practically more difficult than reading financial news (12.18) or The New Republic (11.85).

These findings are unexpected in light of the literature. Consumer contracts appear to be roughly on par with other common sources of information adults consume daily.[10] There are exceptions, and at the tails, both privacy policies and credit card agreements can become complex. But the central mass is similar to the texts adults read voluntarily and habitually. It is also surprising that privacy policies are more complicated than credit transactions, given the relative complexity of the latter over the former. One possible lesson here is that simplification and transactional complexity, while related, are not determinative of each other. It is possible to craft simpler documents for more complex transactions, within some limits of course. It is quite relevant that both types of consumer contracts are regulated. Similarly, privacy policies are subject to very little market pressure—as consumers exhibit significant indifference to them—and so their relative complexity may be attributed to a lack of market discipline.

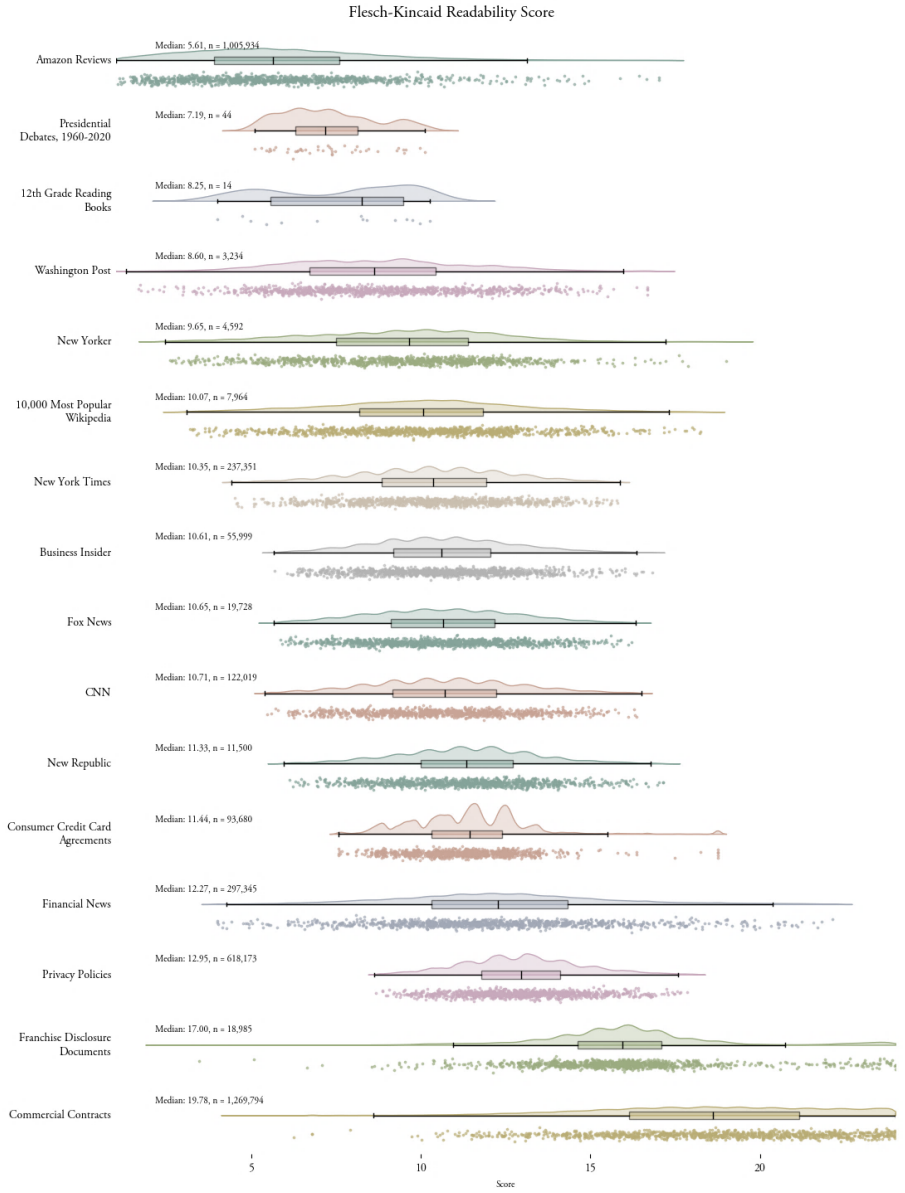
One important remark concerns the benchmark media. As noted, the goal was to amass texts that American adults are known to read daily and voluntarily. Each source of media naturally attracts different audiences, and financial news, for example, has different target audience than Amazon reviews. With that said, credit card agreements fall well within the range of Fox News and CNN, which are broadly accessible sources of news.

While these findings show that consumer contracts are similar to other texts, they also raise one worrisome finding. Franchise disclosure agreements, while nominally applying to franchisee/business-owner, really straddle a grey area between consumer and business law. (HILL & ALBERT, 1979). The reason these disclosure exist in the first place is the concern with exploitation of small business owners by large franchisors.

The following Figure reports these findings on the common Flesch Kincaid measure, adjusted for robustness:

[10] The findings are consistent with the average FRE 42/ FKGL 10.4 that Rustad found in a sample of the thirty-three largest credit card providers (Rustad, 2017)

Figure 6 Robust Flesch-Kincaid Scores by Type of Agreement



The findings largely remain the same, although it is remarkable how differentiated the franchise disclosure agreements are.

1. Contract Length

The following Figure reports the length of the documents reviewed. To preserve the graphical representation, this is represented on a log scale. This means that the differences are much bigger than they are on a normal scale. The first bar has texts with roughly 5 sentences, the last bar features 1,000 sentences. With that in mind, the following figures describes two measures

Figure 7a Log-Scaled Median Number of Sentences by Contract Type

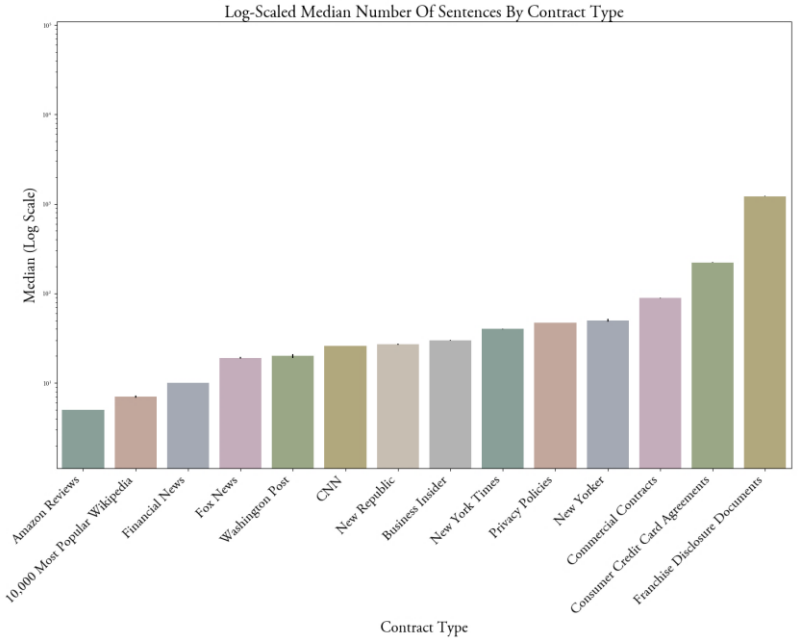
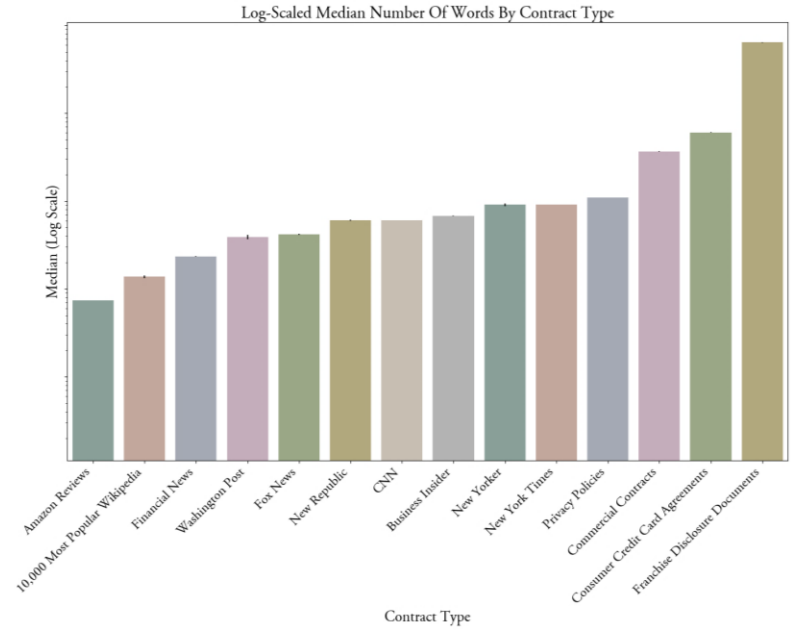


Figure 8b Log-Scaled Median Number of Words by Contract Type



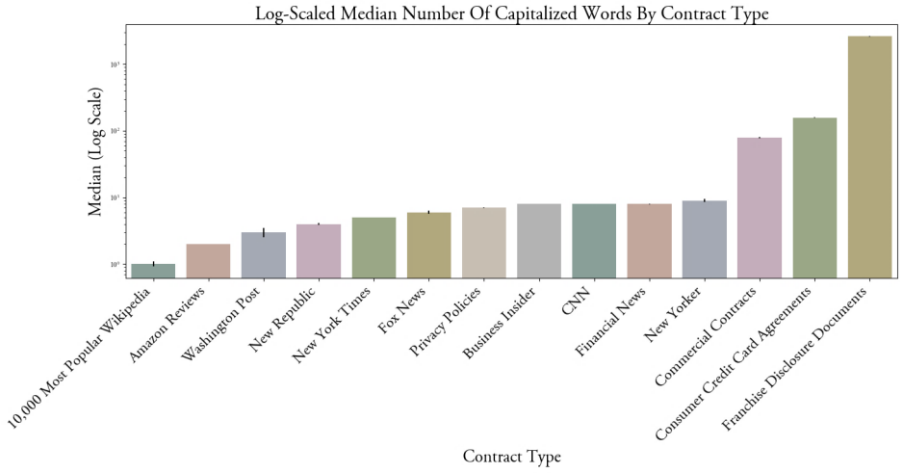
The analysis reveals a few important facts about the length of contracts. As expected, commercial contracts are exceedingly long and pose a significant time investment to read. So are Franchise Disclosure Agreements, which often span hundreds of pages. Even though consumer contracts are at an accessible reading level, they are actually much longer than any of the other texts adults read daily.

As noted above, some argued that plain language invites conciseness. The data does not bear this conclusion. Contracts that would appear readable by the measures used by the movement are quite long. This has two implications: first, it challenges the relevancy of readability tests; second, it highlights the possibility of a sharp trade-off, where improvements to one type of readability can come at the expense of another type, resulting in a net balance.

2. Formatting as measured via ALL-CAPS

The formal readability tests are format agnostic. In reality, however, formatting can have a strong effect on readability and accessibility of texts. There is no standard test for formatting of texts, but one proxy for that, often required by legal codes, is the frequency of using ALL-CAPS (Arbel & Toler, 2020). Research shows that all-caps does not actually improve the readability of texts, and its inclusion can be a signal to its level of formality. The following figure reports, again on a log scale, the median number of capitalized words in the various documents:

Figure 9 Log-Scaled Median Number of Capitalized Words by File Type



The findings show that capitalization is very frequent in commercial contracts and franchise disclosure agreements. Part of it is attributable, no doubt, to their length. But this marks consumer credit agreements in a more negative light—as they are relatively short yet use ample capitalization. One possibility is that this capitalization in credit cards is due to misaligned regulatory design, which encourages this type of formatting, guided by the myth that ALL-CAPS is consumer friendly. The drivers of excessive capitalization in commercial contracts are somewhat harder to decipher. They are certainly not required for the most part, other than the occasional disclaimer. This suggests that capitalization draws not only on legal mandates but also on legal culture. Limiting their usage would require, it would seem, a more holistic approach. At

the same time it is quite surprising that privacy policies contain relatively few of them, perhaps related to new norms in online communications.

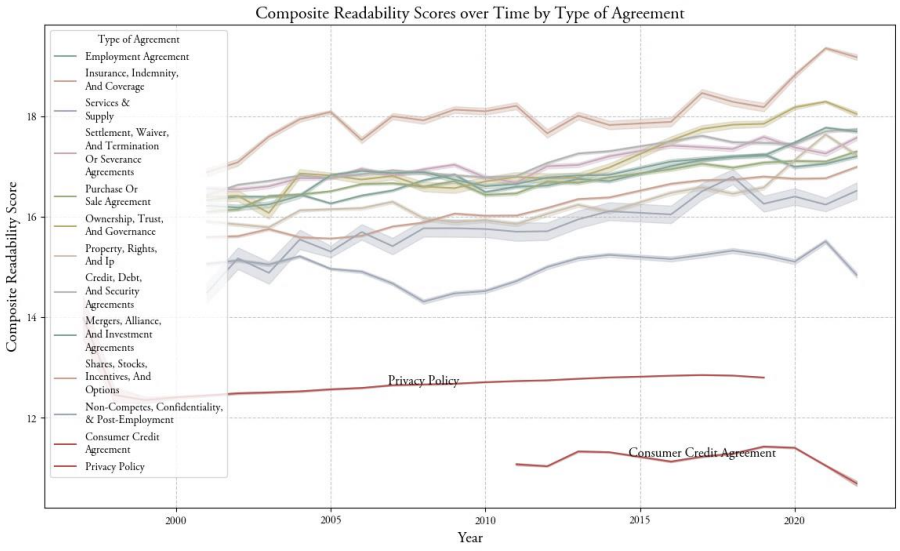
2. The Evolution of Contract Readability Over the Decades

How have contracts evolved over the last two decades? Having a large body of contracts collected over more than two decades allows us to examine trends over time.

1. Readability

The following figure describes the results of the various tests averaged over time, using the weighted average of various readability scores (CRM):

Figure 10 Composite Readability Scores over Time by Type of Agreement



A number of observations emerge from this figure. First, consistent with the cross-section analysis, credit contracts are easier to read than either privacy policies or commercial contracts, and both types of consumer agreements are easier to read than commercial contracts.

Second, commercial contracts are becoming more and more complex over time.[11] The rise is large and significant, signaling a move to greater formality. This trend is interesting because it lends itself to a number of explanations about how technology affects high-end lawyering. While there have been a number of legal changes over the last two decades, none would seem to

[11] One test suggests that all *popular* texts became more sophisticated over time. An examination of 413 stories by the Associated Press from 1962 to 1963 found that the average sentence length was 20 words. The predicted grade level of readability was 6.7. (Catalno, 1990)

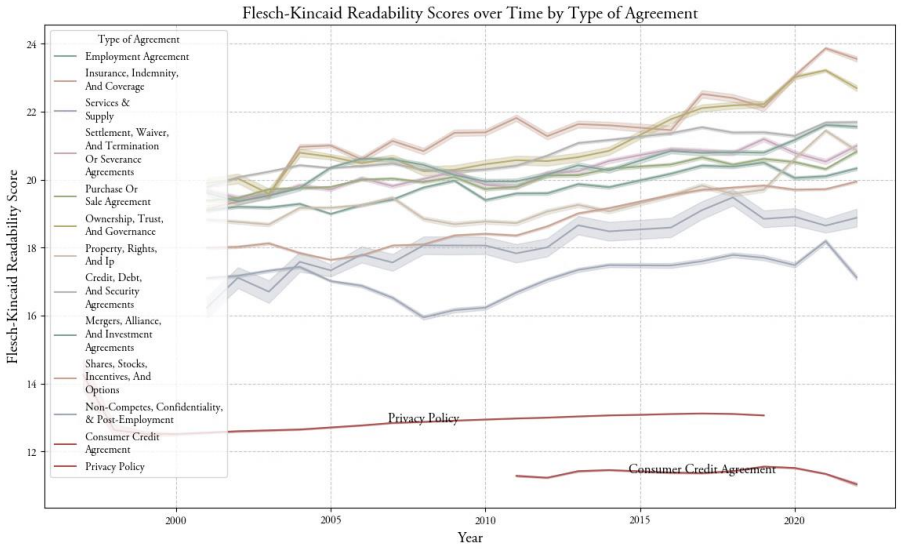
require an overhaul of contract language. Explanations of this trend would depend on non-legal, or at least nondoctrinal, changes. Another important point is that these contracts show that the testing is sensitive to changes over time and therefore can serve as a useful benchmark to the developments in consumer contracts.

Third, after a rocky start, privacy policies have become *easier* to read over time, falling from a fourteenth-grade level to a level somewhat below grade thirteen. I am not aware of any recognition in the literature that this is a possibility. Besides a slow upward creep (which, again, is not discussed in the legal scholarship), there is a large degree of consistency in privacy policies, perhaps due to the high degree of regulation.

Fourth, consumer credit contracts, while easy to read, are remarkably variable. In 2016 we see a marked drop in their difficulty,[12] followed by an upward trend that undoes the change. This is unexpected, because during that time period there was no measure that relaxed plain language requirements, and instead communications from the CFPB generally affirmed its ongoing commitment to plain language (Integrated Mortgage Disclosures Under the Real Estate Settlement Procedures Act (Regulation X) and the Truth In Lending Act(Regulation Z) (12.31.2013)).

The following figure reports on the robust Flesch-Kincaid alone:

Figure 11 Robust Flesch Kincaid Readability Scores



Again, similar trends follow, although we can recognize an interesting uptick in the complexity of consumer credit card agreements post 2020.

[12] In 2015, the collection of credit card agreements was suspended for a year, potentially leading to selection effects or other artifacts that might explain the observed changes in text difficulty. (Bureau of Consumer Financial Protection, 2015)

To better explore changes in the difficulty of commercial contracts, the following table summarizes their readability scores at the beginning and the end of the period for which data exists:

Table 4 Readability Scores by Grade Level, Beginning Period and End of Period

Contract Type	CRM			Flesch Kincaid		
	2001	2022	% Change	2001	2022	% Change
Non-Competes, Confidentiality, & Post-Employment	14.5	16.5	-13.80%	15.9	18.5	-16.40%
Insurance, Indemnity, and Coverage	17	19.2	-13.00%	18.5	25.1	-35.50%
Ownership, Trust, and Governance	16.5	18.1	-9.70%	18.9	23.2	-23.30%
Mergers, Alliance, and Investment Agreements	16.2	17.7	-9.30%	18.9	22.2	-18.70%
Shares, Stocks, Incentives, and Options	15.6	17.2	-9.90%	17.4	19.3	-10.90%
Credit, Debt, and Security Agreements	16.4	17.7	-7.90%	18.8	22.3	-18.60%
Property, Rights, and IP	16	17.2	-7.50%	18.1	20.9	-15.50%
Purchase or Sale Agreement	16.2	17.3	-6.80%	18.3	21	-14.80%
Settlement, Waiver, and Termination or Severance	16.6	17.6	-6.00%	19.3	20.9	-8.30%
Employment Agreement	16.4	17.2	-4.90%	18.7	20	-7.00%
Services & Supply	15.3	15.2	-0.7%	16.5	17.4	-5.50%

The Table summarizes the change in difficulty in the various contract category types over the years. It is policy relevant that non competes have gotten considerably more difficult to read over the last two decades, by around 14%. This is the largest increase among all commercial contracts. That said, almost all commercial contract categories have increased in readability complexity. The only category of commercial contracts that either declined or had a moderate increase is the Services and Supply Agreements category.

2. Length and Reading Time

A common idea in the literature is that contracts are becoming longer over time (Ben-Shahar & Strahilevitz, 2017). The following figures explore this hypothesis, measuring changes in length over time, both in terms of the number of sentences and the number of words in each contract.

Figure 12 Number of Words by Type of Agreement over the Years

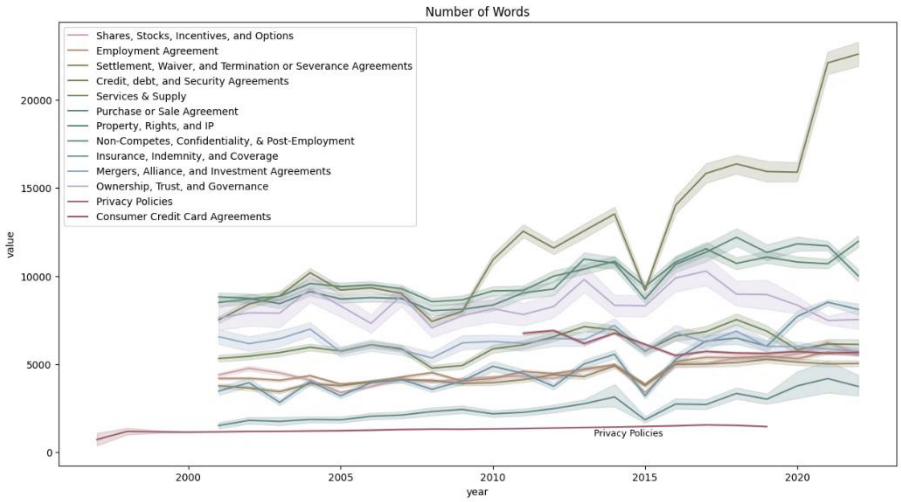
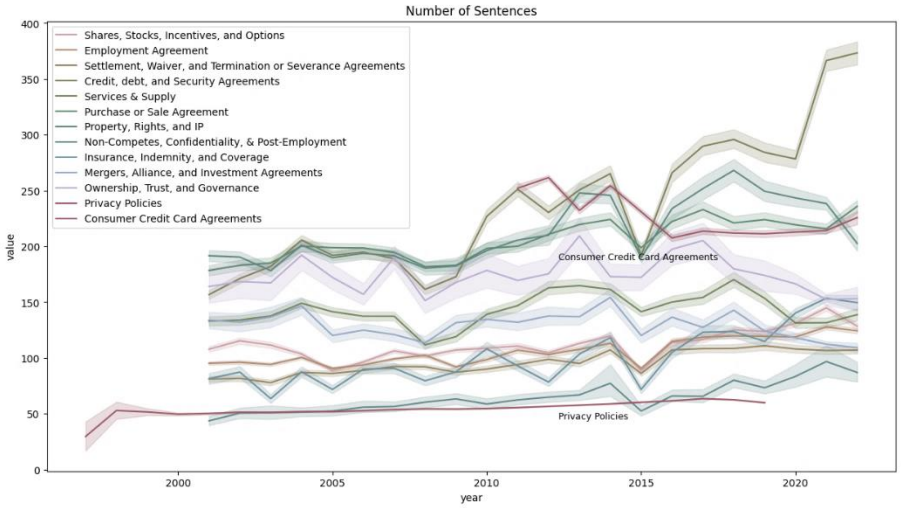


Figure 13 Number of Sentences by Type of Agreement Over the Years



The evidence suggests that consumer contracts have not grown longer over time. Commercial contracts have also remained relatively similar over time, with the unexpected increase in length of settlement and severance agreements. It seems like despite the increase in digitization and ease of reproduction, contract length has remained *relatively* stable.

It is also interesting to see the year to year variations. A copy-and-paste view of contracts would suggest a relatively low level of changes from year to year, but the data suggests that contracts do change. Perhaps some provisions are added and others removed, although the task of analyzing which provisions are added is left for future work.

IV. INTERPRETATION OF FINDINGS AND NORMATIVE IMPLICATIONS

The plain language movement transformed the legal landscape in the USA. The question asked presented in this Article is whether its diagnosis—reading crisis—or its prescription—a broad simplification regimen—were correct. The discussion of the findings here hopes to answer these questions and along the way offer some broader lessons here for consumer protection, theories of regulation, and perhaps even some new questions regarding the ability and responsibility of legal scholarship for its normative prescriptions.

1. Interpretation of Findings

This study, analyzing 2 million contracts and millions of other texts, fundamentally challenges the premises of the plain language movement in contract law. Consumer contracts are no more difficult to read than daily news articles, contradicting decades of assumptions about contract complexity (Benoliel & Becher, 2019; Simkovic & Furth-Matzkin, 2022). This unexpected alignment between contract readability and average adult reading habits undermines the rationale for consumer protection strategies that target abstract linguistic goals.

This is not a rosy conclusion. The analysis adopted the most robust form of readability tests used by regulators, researchers, market actors, and educators. It found that the purported readability crisis is largely non-existent. This suggests that the problems in consumer markets that initially motivated regulatory intervention stem from factors other than linguistic "complexity." Consequently, solutions may need to be more ambitious than mere copyediting of contractual text. This is a challenge for modern contract doctrine, which generally shies away from questions of value and substance, as evidenced by its reluctance to consider the adequacy of consideration (Restatement (Second) of Contracts, 1981; City of Lubbock v. Phillips Petroleum Co., 2000). Yet the findings indicate that a deeper and broader approach may be necessary, one that takes more seriously questions of market structure, consumer choice, and consumer harm (or benefit).

Another important insight for researchers is that despite theoretical incentives for firms to obfuscate, consumer contracts remain relatively readable. Scholars have repeatedly argued that firm power, consumer inattention, consumer myopia, and market discrimination would each separately lead firms to offer the worst terms in their contracts (Arbel & Becher, 2024; Posner & Bebchuk, 2006). This unexpected finding is under-theorized (Hoffman, 2018). It suggests either unexplored market forces favoring clarity or that readability is not the primary tool for firm advantage in consumer contracts. Future research should explore these competing explanations, potentially reshaping our understanding of firm behavior in consumer markets.

Scholars in recent years have started recognizing the importance of consumer documents ex-post, that is, after the consumer has purchased the product. The readability of contracts has implications for dispute resolution,

consumer satisfaction, and vindication of legal rights. The findings here suggest that pathologies in right vindication may be caused by other sources, presumably more structural or substantive, than the readability of consumer agreements.

The analysis thus far has relied on the premise that readability tests are meaningful measures. However, applying Wittgenstein's Ruler principle—which posits that every act of measurement is also a measurement of the metric itself—reveals significant shortcomings in these tests (Wittgenstein, 1978). A central finding of this study is the questionable scientific merit of readability tests championed by the Plain Language Movement.

The critical evaluation of readability metrics yields three key insights. First, readability tests lack reliability and validity, and are easily manipulable, with the same test generating results that are 4.6 years apart on identical text. Second, the weak correlation between tests suggests they fail to capture a unified "readability" construct. Third, a comprehensive review of empirical evidence finds the relationship between these tests and actual consumer outcomes is tenuous at best.

To mitigate these issues, this study employed a robust measure by averaging across and within tests. However, even this approach cannot guarantee sufficient validity for practical application. This presents a significant dilemma for plain language advocates: rejecting these findings on the grounds of test unreliability undermines their claim to scientific rigor and leaves them with no empirical evidence of a readability gap beyond anecdotes and intuition.

While it might be tempting to seek a new readability test geared towards legal texts or consumer contracts, the long history of readability tests suggests such metrics may be counterproductive. Developing "better" tests that account for length, jargon, center-embeddedness, and other linguistic features is relatively straightforward. Indeed, a recent public competition resulted in machine learning models that outperformed standard readability tests. However, replacing Flesch-Kincaid with these tests would likely repeat past mistakes for three main reasons.

First, metrics often perform well in controlled laboratory settings but fail to translate this success to real-world applications. Second, more complex metrics are increasingly susceptible to manipulation, allowing drafters to game the system without genuinely improving readability. Third, and most fundamentally, the pursuit of a universal readability metric is conceptually flawed, as it assumes a single measure can accurately capture readability for all readers across diverse cultural, cognitive, and educational backgrounds.

The plain language movement's goal of reducing document readability to an eighth or sixth-grade level rests on a fundamentally flawed premise: the myth that American adults cannot read at an adult level. This widespread misconception lacks substantive scientific support. Contrary to this belief, Americans rank among the 14 most educated populations in the developed world, with over 51% of adults aged 25-64 possessing tertiary education (OECD, 2022).

The persistence of this myth in scholarly circles raises profound questions about the rigor of legal scholarship and its approach to evidence-based

policymaking. This study's findings expose what Gwern (2009) aptly termed "leprechauns" in academic literature: "deeply-nested chains of citations through ever more obscure sources" that ultimately lead nowhere.

The implications of this mischaracterization are unsettling. Nearly 800 laws have been enacted, many based on this unsubstantiated premise of a readability gap. This misguided effort represents not just a waste of resources in drafting, compliance, and enforcement, but a fundamental miscalibration of regulatory approach. Mandating texts to be written significantly below the actual reading level of the target audience—by five to seven grade levels—potentially undermines the very goals of consumer protection it aims to serve.

While aspirational standard-setting in consumer protection is sometimes laudable, the current approach based on mythical literacy levels is counterproductive. It diverts attention and resources from more pressing issues, particularly the needs of those genuinely struggling with literacy. For this demographic, simplistic solutions like shorter sentences and reduced jargon are inadequate. Instead, they require comprehensive, innovative approaches that may include visual aids, audio materials, official translations, and structural reforms addressing underlying educational and socioeconomic factors.

The last key finding of this study is the divergent evolution of contract readability over time: consumer contract readability has remained remarkably static over decades, while commercial contracts have grown increasingly complex. This unexpected trend presents a significant puzzle for both researchers and policymakers, challenging assumptions about the efficacy of plain language initiatives and market forces in contract drafting.

The stasis in consumer contract readability over the studied period demands careful interpretation. While causal explanations should be approached with caution, three potential hypotheses emerge: ineffectiveness, saturation, and irrelevancy of plain language reforms.

The critical evaluation of the metrics developed in the Article yields three critical insights. First, readability tests lack reliability and validity, and are easily manipulable. In particular, the same readability test can generate results that are 4.6 years apart on the same text. Second, the correlation between the tests is very weak, suggesting that they fail to capture a unified "readability" construct. Third, a comprehensive review of the empirical evidence finds that the relationship of these tests to actual consumer outcomes is tenuous at best. To mitigate these issues, this study employed a robust measure by averaging across and within tests. However, even this approach cannot guarantee sufficient validity for practical application. This presents a significant dilemma for plain language advocates: rejecting these findings on the grounds of test unreliability undermines their claim to scientific rigor and leaves them with no empirical evidence of a readability gap beyond anecdotes and intuition.
It might be tempting at this stage to seek a new readability test, perhaps one geared towards legal texts or consumer contracts. One message from the long history of readability tests is that such metrics may be counterproductive. To be sure, it is fairly easy to develop "better" tests than the antiquated Flesch tests, that will be arbitrarily more sophisticated by accounting for length, jargon,

center-embeddedness, and other linguistic features. Indeed, a recent public competition resulted in machine learning that outcompeted any of the standard readability tests. But to replace Flesch-Kincaid with those tests will be to repeat the mistakes of the past, for three main reasons.

First, metrics often perform well in controlled laboratory settings but fail to translate this success to real-world applications. The gap between lab performance and field validity is a critical issue that proponents of new metrics frequently overlook. Second, more complex metrics are increasingly susceptible to manipulation, allowing drafters to game the system without genuinely improving readability. This vulnerability undermines the very purpose of readability assessment.

Most fundamentally, the pursuit of a universal readability metric is conceptually flawed. The notion that a single measure can accurately capture readability for all readers across diverse cultural, cognitive, and educational backgrounds is untenable. Consumer legal texts vary widely in content and context, and readers bring vastly different experiences and capabilities to their interpretation. This diversity renders the search for a one-size-fits-all readability metric highly suspect.

The plain language movement's goal of reducing document readability to an eighth or sixth-grade level rests on a fundamentally flawed premise: the myth that American adults cannot read at an adult level. This widespread misconception, perpetuated in academic literature and popular media alike, lacks any substantive scientific support. Contrary to this belief, Americans rank among the 14 most educated populations in the developed world, with over 51% of adults aged 25–64 possessing tertiary education (OECD, 2022).

The persistence of this myth in scholarly circles raises profound questions about the rigor of legal scholarship and its approach to evidence-based policymaking. This study's findings expose what Gwern (2009) aptly termed "leprechauns" in legal literature: "deeply-nested chains of citations through ever more obscure sources" that ultimately lead nowhere. The myth of the illiterate American exemplifies this phenomenon, revealing a troubling pattern of accreted misunderstandings and misinterpretations masquerading as fact.

The implications of this mischaracterization should be unsettling. Nearly 800 laws have been enacted, many based on this unsubstantiated premise of a readability gap. This misguided effort represents not just a waste of resources in drafting, compliance, and enforcement, but a fundamental miscalibration of regulatory approach. Mandating texts to be written significantly below the actual reading level of the target audience—by five to seven grade levels—potentially undermines the very goals of consumer protection it aims to serve.

While aspirational standard-setting in consumer protection is sometimes laudable, the current approach based on mythical literacy levels is counterproductive. It diverts attention and resources from more pressing issues, particularly the needs of those genuinely struggling with literacy. For this demographic, simplistic solutions like shorter sentences and reduced jargon are inadequate. Instead, they require comprehensive, innovative approaches that may include visual aids, audio materials, official translations, and structural

reforms addressing underlying educational and socioeconomic factors.
The last key finding of this study is the divergent evolution of contract readability over time: consumer contract readability has remained remarkably static over decades, while commercial contracts have grown increasingly complex. This unexpected trend presents a significant puzzle for both researchers and policymakers, challenging assumptions about the efficacy of plain language initiatives and market forces in contract drafting.
The stasis in consumer contract readability over the studied period demands careful interpretation. While causal explanations should be approached with caution, three potential hypotheses emerge: ineffectiveness, saturation, and irrelevancy of plain language reforms.
The ineffectiveness hypothesis posits that plain language laws have simply failed to impact consumer contracts. This aligns with findings from a comprehensive study of legislative readability, which analyzed every law published between 1951 and 2009, finding no improvement in readability despite concerted efforts. This parallel suggests a systemic challenge in translating readability mandates into practice across different domains of legal text.

The saturation hypothesis proposes that plain language laws have acted as a bulwark, preventing consumer contracts from following the trend of increasing complexity seen in commercial contracts. However, the low variability in consumer contract readability over time casts doubt on this explanation. This stability also contrasts sharply with findings from the insurance sector, where drafters have shown responsiveness to changes in legal norms (Schwarcz, 2020).
The irrelevancy hypothesis suggests that consumer contracts may have already been at an optimal readability level, rendering plain language initiatives superfluous.

Crucially, all three hypotheses point to a sobering conclusion: the failure of plain language law reform. Whether due to ineffectuality, mere preservation of the status quo, or lack of necessity, the implications are equally concerning for the future of consumer protection policy.

So far, the focus has been on pure consumer contracts, but readability also matters for commercial contracts. In the data, commercial contracts and franchise agreements emerge as especially difficult to read. This is neither surprising (Kähler, 2013) and not necessarily worrisome (Lundmark, 2001). When both parties are sophisticated, the ability to craft complex contracts is valuable; the parties can both control and affect their legal design. Even in these contexts, however, the rising difficulty of commercial contracts is surprising. It is possible that the issue is a production cost problem. Claire Hill (2001) once proposed that the length and complexity of commercial contracts are explained by the particularities of their production process. She argued that contractual complexity is rooted in factors such as change costs, cost of legal talent, low return on investment to reducing length, and limits in the internal process itself (Hill, 2001). But while these factors are sensible contenders for explaining complexity, they fail to explain why complexity changes; and they certainly do not explain the rise in complexity in an era when technology makes

simplification easier than ever. There is an unexplained variation of Baumol's cost disease on contract complexity which future work should explore.

To the extent that growing semantic complexity tracks substantive complexity (which is a plausible but untested assumption), the findings carry several important implications for markets. One particular issue is that complex documents are difficult to transfer and adapt. Kevin Davis (2013) argued that such adoption costs stifle contract innovation, leading to market-wide spillover effects. Mitu Gulati and a number of co-authors have proffered adoption cost as a driver of stickiness even in very high-stakes contracts—those relating to sovereign debt (Choi, Gulati, & Posner, 2013). The rise in deal complexity may stifle innovation, even in sophisticated agreements.

Finally, not all commercial contracts are made out of the same material. In particular, franchise agreements involve parties that are not on the same footing as two companies seeking a merger. In franchise markets, many are concerned about the actual sophistication and relative power of franchisees, who are often only a step removed from ordinary consumers (Levine, 2022). Indeed, this is the reason that disclosure requirements exist in the first place (Federal Trade Commission [FTC], 1978; FTC, 2007). The findings here suggest that the extensive disclosure requirements may be ineffectual, given the level of difficulty of disclosed materials.

2. Policy Prescriptions

The findings naturally leads to the normative question: *what now*? How should the field of consumer protection in contract law evolve in light of the apparent shortcomings of plain language reform? This section outlines three key policy implications that should inform future lawmaking and scholarly discourse.

2.1. Reassessing the Value of Plain Language Reform

The first implication stems from a sobering realization: not every well-intentioned policy intervention yields its intended benefits. The plain language movement, despite its laudable goals, has resulted in over 800 laws whose costs may outweigh their modest and uncertain benefits. These costs are both direct (e.g., resources spent on drafting, implementing, and enforcing plain language laws) and indirect (e.g., opportunity costs of not pursuing alternative reforms).

For instance, the resources devoted to simplifying contract language could have been directed towards addressing more fundamental issues in consumer markets, such as improving market competition to give consumers more meaningful choices; addressing information asymmetries through targeted disclosure requirements; strengthening enforcement mechanisms against unfair or deceptive practices; or removing regulatory barriers to entry to markets.

Perhaps there are still areas where targeted plain language reform could help, but given the findings of this study, the burden of proof now shifts to proponents of plain language reform. Future proposals must demonstrate not only that reform is needed, but also that it is likely to be effective and worth the associated costs. They should specifically identify specific types of contracts or market segments where linguistic simplification could yield substantial benefits;

develop more robust and valid measures of contract readability; and finally establish a more accurate baseline of American adults' reading capabilities and habits *in that market segment.*

2.2. Shifting Focus to Consumer-Side Tools: The Promise of Smart Readers

The second implication challenges the premise that pro-consumer drafting should be primarily the responsibility of firms. The problem is not so much the cost readability reform imposes on firms: this is a fairly small part of the cost of doing business. The problem is that this structure is not incentive compatible, manipulable, and often counterproductive. Firms do not have the right incentives to make certain legal (as opposed to marketing) materials accessible. By imposing a duty for them to abide by manipulable tests, firms are actually given more leeway to obfuscate than they would under broader standards.

The deeper challenge is that the average reader may well be misdirection. Readers have broadly different linguistic, cognitive, and sociocultural needs. Marking the hypothetical average reader as the gold standard of contract drafting redirects attention from the diverse to the modal. An effective drafting strategy should aim at the tails.

Of course, it is hard to do so, because that would require that the same document will be drafted with very different audiences in mind. The blanket is perhaps too short to cover non native speakers, people with short attention spans, or people with low eyesight in a single document. But the single document is a fallacy.

This is where consumer-side tools, particularly AI-powered "smart readers," come into play. Recent advances in machine learning have led to a breakthrough in language models, enabling the development of smart readers with several core capabilities. These tools can simplify complex legal language, personalize the presentation of contractual information, interpret terms and their implications, and even benchmark contracts based on their quality and consumer-friendliness.

A smart reader can be calibrated to present the same document to a 15-year-old, to a person who appreciates bullet point summaries, to one who likes visual presentation, or a person preferring Spanish to English. Firms cannot do this, let alone do so reliably. But smart readers can.

The potential benefits of smart readers are significant. They can overcome traditional information barriers that have long plagued consumer contracts, empower consumers with personalized, on-demand contract analysis, reduce reliance on intermediaries for interpretation, and potentially close gaps in access to justice by making legal information more accessible.

However, the implementation and regulation of smart readers also present challenges. These include ensuring the accuracy and reliability of AI-generated interpretations, addressing potential biases in the underlying language models, protecting user privacy and data security, and determining liability in cases of misinterpretation or malfunction.

Despite these challenges, pivoting regulatory attention to the broad assimilation of smart readers should be a primary goal for those who care about access rights. This shift in focus from firm-side obligations to consumer-side empowerment tools represents a paradigm change in how we approach contract readability and consumer protection.

To facilitate this transition, policymakers should consider several key steps. A central policy effort should be directed at the development of standards and certification processes for smart reader applications. Creating legal frameworks to address liability and consumer protection issues related to AI-powered legal tools will be necessary to ensure trust in these systems. Moreover, promoting accessibility will be essential to ensure consumers can effectively utilize smart readers. Finally, policymakers might consider incentivizing the development and adoption of smart reader technology through various economic mechanisms.

By embracing smart reader technology, we can move beyond the limitations of one-size-fits-all plain language requirements and towards a more flexible, personalized approach to contract comprehension. This approach not only addresses the diverse needs of consumers but also aligns with the rapidly evolving technological landscape of the 21st century. It offers a promising path forward in our ongoing efforts to enhance consumer protection and contract accessibility in an increasingly complex legal and technological environment.

2.3. Encouraging Innovative Interventions Beyond Readability

While this study challenges the efficacy of traditional plain language approaches, it should not be interpreted as a license for firms to obfuscate or complicate contracts. Instead, these results should spur scholars and policymakers to explore innovative interventions that move beyond simplistic notions of readability. As a close example, Lauren Willis wrote an important article on the failure of financial literacy (Willis, 2021). Her point was not silent acceptance of consumer lack of sophistication, but rather, that new paradigms are necessary—and subsequent work has indeed explored such alternatives.

Future policy and scholarly work would therefore benefit from focus on other types of potential interventions. First, if problems persist, perhaps substantive regulation will be unavoidable. Rather than fixating on linguistic form, policymakers should scrutinize the actual content and fairness of contract terms. This approach addresses the root causes of consumer harm more directly than linguistic simplification.

Second, market-based solutions offer untapped potential. Researchers should investigate ways to harness competitive forces that naturally incentivize firms to offer more consumer-friendly contracts. This might involve creating mechanisms for easy comparison of contract terms across providers or developing consumer-friendly ratings systems for contract fairness.

Finally, technological interventions beyond smart readers deserve exploration. For instance, AI shopping assistants can help consumers make an all-things-considered informed decisions and bridge some of the sophistication gap between them and sellers.

CONCLUSION

The narrative of the rise of the plain language movement is a story of immense success and immense failure. The movement led the largest consumer revolution of the last generation, it spread at the state level and the federal level, it spans contracts, regulations, and laws. Yet, for all of its zeal and influence, it is doubtful that any of its promises were ever realized. The current study reviewed with great attention the foundational tenets of this movement. Every single premise failed scrutiny. Americans do not really read at the eighth-grade level. Simplification is not a silver bullet to improving consumer outcomes and only seem to carry marginal positive effects. It is doubtful that readability can be easily quantified, and even if readability can be measured in theory, the popular metrics used today are unreliable and manipulable. And most centrally, it is simply not the case that consumer contracts are beyond the grasp of most Americans.

The primary normative implication is the need for regulators to focus more on questions on substance, market structure, and choice in designing consumer regulation. In previous work, I borrowed money from payday lenders (Arbel, 2020). The least expected part of the experience was how *nice* it was. The clerk treated the clients with patience. She explained the terms slowly and clearly. Soft jazz music played in the background. A regulatory mandated poster disclosed, in very large print, the terms of the transaction. I am hard pressed to think of better financial experiences I have had elsewhere. And yet, the deal was rotten. I was quoted 600% APR, about 30 times the rate on my already expensive credit card.

The point of this digression is that readability is neither a necessary nor sufficient condition to resolving market pathologies. When the substance of agreements is fair and standard, their readability does not matter, and when terms are exploitative and problematic, readability does little to cure them (Oman, 2023). Doug Baird explains that much of the scholarly focus on boilerplate is really about discomfort with problematic market structures that are projected on forms as a scapegoat of sorts (Baird, 2006).

Another implication of these findings is the need, in consumer protection law, to focus more on vulnerable individuals. There are a great idiosyncratic many who struggle in transactional contexts for a variety of issues; some do not speak English well, others may be neurodivergent, yet others may need visuals rather than text. Especially important for the future of consumer protection is the integration of consumer-assistive AI technology. Today, when smart reader technology is available, such special needs can be effectively addressed. Instead of relying on the goodwill of firms to simplify contracts, regulators should think of mechanisms like APIs that would allow consumers to easily interact with their agreements.

Finally, the analysis also raised some broader questions on legal movements and how they might perpetuate myths in the name of higher calling. The plain language movement has had the best of intentions; it still has them. But in pursuing them it failed to properly underwrite its promises. After half a century of efforts, it is only natural to ask whether we live in a world where

consumers read more contracts, understand them better, and are making informed decisions. I believe the answer to this question is plain.

1. BIBLIOGRAPHY

Agency for Healthcare Research & Quality. "Saying It Clearly." *TalkingQuality Project*, 2008.

Aiken, Ray J. "Let's Not Oversimplify Legal Language." *Rocky Mountain Law Review* 32 (1960): 358–72.

Amos, Ryan, Gunes Acar, Elena Lucherini, Mihir Kshirsagar, Arvind Narayanan, and Jonathan Mayer. "Privacy Policies Over Time: Curation and Analysis of a Million-Document Dataset." Working paper, Department of Computer Science, Cornell University, Ithaca, NY, 2020.

Ante, Lennart. "The Relationship Between Readability and Scientific Impact: Evidence from Emerging Technology Discourses." *Journal of Informetrics* 16 (2022): 101275.

Arbel, Yonathan A. "Payday." *Washington University Law Review* 98, no. 1 (2020): 1–59.

Arbel, Yonathan A., and Samuel Becher. "Contracts in the Age of Smart Readers." *George Washington Law Review* 90, no. 1 (2022): 83–146.

---. "How Smart Are Smart Readers? LLMs and the Future of the No-Reading Problem." *The Cambridge Handbook on Emerging Issues at the Intersection of Commercial Law and Technology*, edited by Nancy Kim and Stacey-Ann Elvy, Cambridge University Press, 2024.

Arbel, Yonathan A., and David Hoffman. "Generative Interpretation." *New York University Law Review*, forthcoming 2025.

Arbel, Yonathan A., and Roy Shapira. "Theory of the Nudnik: The Future of Consumer Activism and What We Can Do to Stop It." *Vanderbilt Law Review* 73, no. 4 (2020): 929–1002.

Arbel, Yonathan A., and Andrew Toler. "ALL-CAPS." *Journal of Empirical Legal Studies* 17, no. 4 (2020): 862–96.

Ardoin, Scott P. "Accuracy of Readability Estimates' Predictions of CBM Performance." *School Psychology Quarterly* 20, no. 1 (2005): 1–22.

Asprey, Michele. *Plain Language for Lawyers*. 2nd ed., The Federation Press, 1996.

Ayres, Ian, and Alan Schwartz. "The No-Reading Problem in Consumer Contract Law." *Stanford Law Review* 66, no. 3 (2014): 545–610.

Ayres, Ian, and Eyal Zamir. "A Theory of Mandatory Rules: Typology, Policy, and Design." *Texas Law Review* 99, no. 2 (2020): 283.

Baird, Douglas G. "The Boilerplate Puzzle." *Michigan Law Review* 104, no. 5 (2006): 933–52.

Bakos, Yannis, Florencia Marotta-Wurgler, and David R. Trossen. "Does Anyone Read the Fine Print? Consumer Attention to Standard-Form Contracts." *Journal of Legal Studies* 43, no. 1 (2014): 1–35.

Baldwin, Janet, et al. "The Literacy Proficiencies of GED Examinees: Results from the GED-NALS Comparison Study." *American Council on Education and Educational Testing Service*, 45–6. 1995. .

Bar-Gill, Oren. "Seduction by Plastic." *Northwestern University Law Review* 98, no. 4 (2004): 1373–434.

Barnes, Jonathan M. "Tailored Jury Instructions: Writing Instructions That Match a Specific Jury's Reading Level." *Mississippi Law Journal* 87, no. 2 (2017): 193–252.

Bartlett, Robert P., and Victoria C. Plaut. "Blind Consent? A Social Psychological Investigation of Non-Readership of Click-Through Agreements." *Law and Human Behavior* 35 (2011): 293–309.

Bebchuk, Lucian A., and Richard A. Posner. "One-Sided Contracts in Competitive Consumer Markets." *Michigan Law Review* 104, no. 5 (2006): 827–36.

Becher, Samuel I., and Esther Unger-Aviram. "The Law of Standard Form Contracts: Misguided Intuitions and Suggestions for Reconstruction." *DePaul Business and Commercial Law Journal* 8, no. 3 (2010): 199–227.

Becher, Samuel I., and Uri Benoliel. "Law in Books and Law in Action: The Readability of Privacy Policies and the GDPR." *Consumer Law & Economics* 9 (2021): 179–204.

Becher, Samuel I., and Tal Z. Zarsky. "Minding the Gap." *Connecticut Law Review* 51 (2019): 69–120.

Ben-Shahar, Omri, and Carl E. Schneider. *More Than You Wanted to Know: The Failure of Mandated Disclosure*. Princeton University Press, 2014.

Ben-Shahar, Omri, and Lior Strahilevitz, *Interpreting Contracts via Surveys and Experiments*, Coase-Sandor Working Paper Series in Law and Economics, No. 791 (2017)

Ben-Shahar, Omri. "The Myth of the 'Opportunity to Read' in Contract Law." *European Review of Contract Law* 5 (2009): 1–20.

Benoliel, Uri, and Shmuel I. Becher. "The Duty to Read the Unreadable." *Boston College Law Review* 60, no. 8 (2019): 2255–96.

Benoliel, Uri, and Xu (Vivian) Zheng. "Are Disclosures Readable? An Empirical Test." *Alabama Law Review* 70, no. 1 (2018): 237–68.

Benson, Robert W. "The End of Legalese: The Game is Over." *New York University Review of Law & Social Change* 13, no. 3 (1984): 519–73.

Bentham, Jeremy. *The Works of Jeremy Bentham*. Edited by John Bowring, vol. 5, William Tait, 1843.

Black, Bernard. "A Model Plain Language Law." *Stanford Law Review* 33, no. 2 (1981): 255–300.

---. "The Rise of Plain Language Laws." *University of Miami Law Review* 76 (2022): 447–514.

Blasie, Michael. "Regulating Plain Language." *Wisconsin Law Review* 2023 (2023): 687–764.

---. "Credit Card Agreement Database." .

---. "Policy Statement on Abusive Acts or Practices." April 3, 2023.

---. "Temporary Suspension of Credit Card Agreement Submissions." *Federal Register* 80, no. 74 (April 17, 2015): 21256–59.

Bureau of Consumer Financial Protection. *The Consumer Credit Card Market*. Washington, DC: Bureau of Consumer Financial Protection, September 2021. .

Cadwell, Patrick, Alessandra Rossetti, and Sharon O'Brien. "The Terms and Conditions Came Back to Bite." *HCI International 2020 – Late Breaking Papers: Interaction, Knowledge and Social Media*, edited by Constantine Stephanidis, Margherita Antona, and Stavroula Ntoa, 699–704. Cham, Switzerland: Springer, 2020.

Calamari, John D., and Joseph M. Perillo. *Calamari and Perillo on Contracts*. 6th ed., West Academic Publishing, 2009.

Chall, Jeanne S., and Edgar Dale. *Readability Revisited: The New Dale-Chall Readability Formula*. Brookline Books, 1995.

Charrow, Robert P., and Veda R. Charrow. "Making Legal Language Understandable: A Psycholinguistic Study of Jury Instructions." *Columbia Law Review* 79, no. 7 (1979): 1306–74.

---. "Assessing Readability Formulas: A Comparison of Readability Formula Performance on the Classification of Simplified Texts." *Research in Social and Administrative Pharmacy* 9 (2013): 503–13.

---. "Readability Assessment Tool for English Tests." Working Paper No. 6190, Department of Applied Linguistics & ESL, Georgia State University, Atlanta, GA, 2021.

Choi, Joon Suh, and Scott A. Crossley. "Advances in Readability Research: A New Readability Web App for English." *Proceedings of the 2022 IEEE International Conference on Advanced Learning Technologies (ICALT)*, 1–5. New York: IEEE, 2022.

Choi, Stephen J., G. Mitu Gulati, and Eric A. Posner. "The Dynamics of Contract Evolution." *New York University Law Review* 88, no. 1 (2013): 1–50.

Ciocchetti, Corey. "The Privacy Matrix." *Journal of Technology Law & Policy* 12, no. 2 (2007): 245–329.

Conn. Gen. Stat. Ann. § 42-152 (West 2012).

Coughlin, Christine. "E-Consent: Can Informed Consent Be Just a Click Away?" *Wake Forest Law Review* 50, no. 2 (2015): 381–98.

Crossley, Scott A., Kristopher Kyle, and Danielle S. McNamara. "Moving Beyond Classic Readability Formulas: New Methods and New Models." *Journal of Research in Reading* 42, no. 3–4 (2019): 541–61.

Crossley, Scott, Eleni Ilkou, Achilles Kameas, and Panagiotis Bamidis. "Assessing Readability Formulas in the Wild." *Polyphonic Construction of Smart Learning Ecosystems*, edited by Eleni Ilkou, Achilles Kameas, and Panagiotis Bamidis, 91–103. Cham, Switzerland: Springer, 2022.

Crump, David. "Against Plain English: The Case for a Functional Approach to Legal Document Preparation." *Rutgers Law Journal* 33, no. 3 (2002): 713–44.

Danish Shakeel, M., & Paul E. Peterson. "A Half Century of Progress in U.S. Student Achievement: Agency and Flynn Effects, Ethnic and SES Differences." *Educational Psychology Review* 34 (2022): 1255.

Dapcevich, Madison. "Do More Than Half of Americans Read Below 6th-Grade Level?" *Snopes*, August 2, 2022.

David v. M.L.G. Corp., 712 P.2d 985 (Colo. 1986).

Davis, Jeffrey. "Protecting Consumers from Overdisclosure and Gobbledygook: An Empirical Look at the Simplification of Consumer-Credit Contracts." *Virginia Law Review* 63, no. 6 (1977): 841–920.

Davis, Kevin E. "Contracts as Technology." *New York University Law Review* 88, no. 1 (2013): 83–127.

Del. Admin. Code § 605-4.0 (2023).

Desmet, Pieter, Mark Van Dam, and Willem H. Van Boom. "If It's Easy to Read, It's Easy to Claim" – The Effect of the Readability of Insurance Contracts on Consumer Expectations and Conflict Behaviour. *Journal of Consumer Policy* 39 (2016): 187–209.

Dixon Weaver, Jessica. "Racial Myopia in [Family] Law." *Yale Law Journal Forum* 132 (2023): 1086–1106.

Doak, Cecelia Conrath, et al. *Teaching Patients with Low Literacy Skills*. 2nd ed., 1996, 2–3.

Dumas, Bethany K. "Jury Trials: Lay Jurors, Pattern Jury Instructions, and Comprehension Issues." *Tennessee Law Review* 67, no. 3 (2000): 701–42.

Dyer, Charles R., Joan E. Fairbanks, M. Lee Greener, Kimberly Barker, Lois Skeen, and Josefina Cerrillo-Ramirez. "Improving Access to Justice: Plain Language Family Law Court Forms in Washington State." *Seattle Journal for Social Justice* 11, no. 3 (2013): 1065–1111.

Elvy, Stacy-Ann. "Commodifying Consumer Data in the Era of the Internet of Things." *Boston College Law Review* 59, no. 2 (2018): 423–522.

Epstein, Richard A. "Contract, Not Regulation: UCITA and High-Tech Consumers Meet Their Consumer Protection Critics." *Consumer Protection in the Age of the 'Information Economy'*, edited by Jane K. Winn, 205–34. Carolina Academic Press, 2006.

Fair Debt Collection Practices Act, 15 U.S.C. §§ 1692–1692 (2018).

Farkash, Yuval, and Eyal Zamir. "Standard Form Contracts: Empirical Studies, Normative Implications, and the Fragmentation of Legal Scholarship." *Journal of Review of Legal Studies* 12 (2015): 137–67.

Federal Communications Commission. "It's Time to Reexamine the FCC's Kid Vid Requirements." *Federal Communications Commission Blog*, January 26, 2018. .

Federal Plain Language Guidelines. U.S. General Services Administration. .

Federal Reserve Bank of Atlanta. "The 2020 Survey of Consumer Payment Choice: Summary Results." Federal Reserve Bank of Atlanta, April 28, 2021.

Feld, Bernard, Jr. "Empirical Test Proves Clarity Adds Readers." *Editor & Publisher*, April 17, 1948, 29.

Felsenfeld, Carl. "The Plain English Movement in the United States." *Canadian Business Law Journal* 6 (1981–1982): 408–21.

---. *Marks of Readable Style: A Study in Adult Education*. Bureau of Publications, Teachers College, Columbia University, 1943.

Fitriya, Luluk Anjar, Niki Kunang Prasasti, and Indah Werdiningsih. "A Review of Data Compression Techniques." *International Journal of Applied Engineering Research* 12, no. 19 (2017): 8956–63.

Fitzsimmons, Paul R., B. D. Michael, J. L. Hulley, and G. O. Scott. "A Readability Assessment of Online Parkinson's Disease Information." *Journal of the Royal College of Physicians of Edinburgh* 40, no. 4 (2010): 292–97.

Fla. Stat. Ann. § 501.71 (West 2023).

Flesch, Rudolf F. *How to Write Plain English: A Book for Lawyers and Consumers*. Harper & Row, 1979.

Foster, Kevin, Claire Greene, and Joanna Stavins. "The 2020 Survey of Consumer Payment Choice: Summary Results." Federal Reserve Bank of Atlanta, April 28, 2021.

Friman, Michael S. "Plain English Statutes: Long Overdue or Underdone?" *Loyola Consumer Law Reporter* 7, no. 3 (1995): 103–12.

Fry, Edward B. "Reading Formulas: Maligned but Valid." *Journal of Reading* 32, no. 4 (1989): 292–97.

Furth-Matzkin, Meirav, and Michael Simkovic. "Proportional Contracts." *Iowa Law Review* 107 (2022): 238–308.

Gallacher, Ian. "'When Numbers Get Serious': A Study of Plain English Usage in Briefs Filed Before the New York Court of Appeals." *Suffolk University Law Review* 46, no. 2 (2013): 451–504.

Gariulo, Albert F., and Paul T. Hill. *Contracts: The Move to Plain Language*. 37. Practicing Law Institute, 1979.

Garner, Bryan A. *Legal Writing in Plain English: A Text with Exercises*. 2nd ed., University of Chicago Press, 2013.

Gelles-Watnick, Risa, and Andrew Perrin. "Who Doesn't Read Books in America?" *Pew Research Center*, September 21, 2021. .

General Data Protection Regulation (GDPR), art. 4(1).

Gibson, Edward, Eric Martínez, and Francis Mollica. "Poor Writing, Not Specialized Concepts, Drives Processing Difficulty in Legal Language." *Cognition* 224 (2022): 104857.

Gillette, Clayton P. "Rolling Contracts as an Agency Problem." *Wisconsin Law Review* 2004, no. 2 (2004): 679–721.

Gopen, George D. "The State of Legal Writing: Res Ipsa Loquitur." *Michigan Law Review* 86, no. 2 (1987): 333–80.

Government Accountability Office. *Credit Cards: Increased Complexity in Rates and Fees Heightens Need for More Effective Disclosures to Consumers*. U.S. Government Printing Office, 2006.

Green, Justin P. "The Consumer-Redistributive Stance: A Perspective on Restoring Balance to Transactions Involving Consumer Standard-Form Contracts." *Akron Law Review* 46, no. 3 (2013): 551–78.

Guthrie, James W., and Matthew G. Springer. "A Nation at Risk Revisited: Did 'Wrong' Reasoning Result in 'Right' Results? At What Cost?" *Peabody Journal of Education* 79, no. 1 (2004): 7–35.

H.R. 946, Plain Writing Act of 2010, 111th Cong. (2010).

Hathaway, George. "An Overview of the Plain English Movement for Lawyers . . . Ten Years Later." *Michigan Bar Journal* 73, no. 1 (1994): 26–29.

Herrine, Luke. "The Folklore of Unfairness." *New York University Law Review* 96, no. 2 (2021): 431–511.

Heydari, Pooneh, and A. Mehdi Riazi. "Readability of Texts: Human Evaluation Versus Computer Index." *Mediterranean Journal of Social Sciences* 3, no. 1 (2012): 177–90.

---. "Why Contracts Are Written in 'Legalese'." *Chicago-Kent Law Review* 77, no. 1 (2001): 59–85.

Heydari, Pooneh. "The Validity of Some Popular Readability Formulas." *Mediterranean Journal of Social Sciences* 3, no. 2 (2012): 423–35.

---. "On-Line Consumer Standard-Form Contracting Practices: A Survey and Discussion of Legal Implications." *Cornell Legal Studies Research Paper* No. 05-012. Cornell Law School, 2005. .

Hill v. Gateway 2000, Inc., 105 F.3d 1147 (7th Cir. 1997)

Hill, Claire A. "Bargaining in the Shadow of the Lawsuit: A Social Norms Theory of Incomplete Contracts." *Delaware Journal of Corporate Law* 34, no. 1 (2009): 191–220.

Hillman, Robert A. "Online Boilerplate: Would Mandatory Website Disclosure of E-Standard Terms Backfire?" *Michigan Law Review* 104, no. 5 (2006): 837–56.

Hillman, Robert A., and Ibrahim Barakat. "Warranties and Disclaimers in the Electronic Age." *Yale Journal of Law and Technology* 11 (2009): 1–67.

Hillman, Robert A., and Jeffrey J. Rachlinski. "Standard-Form Contracting in the Electronic Age." *New York University Law Review* 77, no. 2 (2002): 429–95.

Hoffman, David A. "Relational Contracts of Adhesion." *University of Chicago Law Review* 85, no. 6 (2018): 1395–1459.

Hoffman, David. "What If We'd Already Revolutionized Contract Law But No One Knew it?" *Journal of Online Technology & Legal Scholarship (JOTWELL)*, March 8, 2022. .

Hulley, J. L., Paul R. Fitzsimmons, B. D. Michael, and G. O. Scott. "A Readability Assessment of Online Parkinson's Disease Information." *Journal of the Royal College of Physicians of Edinburgh* 40, no. 4 (2010): 292–97.

Ilkou, Eleni, Achilles Kameas, and Panagiotis Bamidis. *Polyphonic Construction of Smart Learning Ecosystems*. Springer, 2022.

Jones, Jeffrey M. "Americans Reading Fewer Books Than In Past." *Gallup*, January 10, 2022. .

Kameas, Achilles, Eleni Ilkou, and Panagiotis Bamidis. *Polyphonic Construction of Smart Learning Ecosystems*. Springer, 2022.

Katz, Avery. "Your Terms or Mine? The Duty to Read the Fine Print in Contracts." *RAND Journal of Economics* 21, no. 4 (1990): 518.

Keys, Benjamin J., and Jialan Wang. "Minimum Payments and Debt Paydown in Consumer Credit Cards." *National Bureau of Economic Research Working Paper* No. 22742, 2016.

Kimble, Joseph. "Answering the Critics of Plain Language." *Scribes Journal of Legal Writing* 5 (1995): 51.

Kolt, Noam. "Predicting Consumer Contracts." *Berkeley Technology Law Journal* 37, no. 1 (2022): 71.

Kornhauser, Lewis A. "Unconscionability in Standard Forms." *California Law Review* 64, no. 5 (1976): 1151.

Korobkin, Russell. "Bounded Rationality, Standard Form Contracts, and Unconscionability." *University of Chicago Law Review* 70, no. 4 (2003): 1203.

Kraczon, Caroline, Katherine Ireland, and Tim Samples. "The Law and Linguistics of Social Platform Terms-of-Use." *Berkeley Technology Law Journal* 39 (2024): 47–106.

Kuczmarski, Marie F., et al. "Health Literacy and Education Predict Nutrient Quality of Diet of Socioeconomically Diverse, Urban Adults." *Journal of Epidemiology and Preventive Medicine* 2, no. 1 (2016).

Kähler, Lorenz. "Contract-Management Duties as a New Regulatory Device." *Law and Contemporary Problems* 76 (2013): 89–128.

La. Code Civ. Ann. art. 2046 (2011).

Lemley, Mark A. "The Benefit of the Bargain." *Wisconsin Law Review* 2023 (2023): 237.

Levine, Samuel. "Holding Franchisors Accountable for Illegal Practices." *FTC Blog*, August 3, 2022.

Listokin, Yair. "The Meaning of Contractual Silence: A Field Experiment." *Journal of Legal Analysis* 2, no. 2 (2010): 397.

Logue, Kyle D., Daniel Schwarcz, and Brenda J. Cude. "The Value and Reality of Transparent Consumer Insurance Contracts." Working paper, October 27, 2022, 16–17.

Lundmark, Thomas. "Verbose Contracts." *American Journal of Comparative Law* 49 (2001): 121–21.

Macro International Inc. *Design and Testing of Effective Truth in Lending Disclosures*. Calverton, MD: Macro International Inc., 2007. .

Magnuson-Moss Warranty Act, 15 U.S.C. §§ 2302(a) (1975).

Marotta-Wurgler, Florencia, and Robert Taylor. "Set in Stone? Change and Innovation in Consumer Standard-Form Contracts." *New York University Law Review* 88, no. 1 (2013): 240.

Marotta-Wurgler, Florencia. "Will Increased Disclosure Help? Evaluating the Recommendations of the ALI's 'Principles of the Law of Software Contracts." *University of Chicago Law Review* 78, no. 1 (2011): 165.

Martínez Silvagnoli, Leia, et al. "Optimizing Readability and Format of Plain Language Summaries for Medical Research Articles: Cross-sectional Survey." *Journal of Medical Internet Research* 24, no. 1 (2022).

Masson, Michael E. J., and Mary Anne Waldron. "Comprehension of Legal Contracts by Non-Experts: Effectiveness of Plain Language Redrafting." *Applied Cognitive Psychology* 8, no. 1 (1994): 67.

Matwyshyn, Andrea M. "Resilience: Building Better Users and Fair Trade Practices in Information." *Federal Communications Law Journal* 63, no. 3 (2011): 391.

McCarty, Kendall, and Jayne Rogers. "Inpatient Asthma Education Program." *Pediatric Nursing* 38, no. 4 (2012): 257.

McClure, Glenda M. "Readability Formulas: Useful or Useless?" *IEEE Transactions on Professional Communication* 30, no. 1 (1987): 95.

Mellinkoff, David. *The Language of the Law*. Little, Brown and Company, 1963.

MetaMetrics. "Lexile Text Analyzer." .

Mich. House Legislative Analysis Section. "Analysis of H.B. 4028." December 4, 1997.

Minn. Stat. Ann. § 325.

Miyazaki, Susumu. "Should Japan Adopt a Plain Language Rule?" *Minnesota Journal of Global Trade* 13, no. 1 (2004): 1.

Mont. Code Ann. § 30-14-1103 (West 1985).

More, Thomas. *Utopia*. Penguin Classics, 2021.

Mowat, Christine. "Exploring Meanings of Plain Language." *Clarity* 52 (2004): 42.

N.J. Stat. Ann. § 56:12-10 (West 2007).

---. "We'll See You in … Court! The Lack of Arbitration Clauses in International Commercial Contracts." *International Review of Law and Economics* 58 (2019): 6.

N.Y. Gen. Oblig. Law § 5-702 (West 1978).

Nasaa Electronic Filing Directory. North American Securities Administrators Association. .

Nyarko, Julian. "Stickiness and Incomplete Contracts." *University of Chicago Law Review* 88, no. 1 (2021): 1.

---. "Intent to Create Legal Relations and the Nature of Contractual Consent." *Research Handbook of the Philosophy of Contract Law*, edited by Nathan B. Oman. Edward Elgar Publishing, forthcoming 2023.

OECD Skills Surveys: PIAAC. .

OECD. *Education at a Glance 2022: OECD Indicators*. 2022.

Oman, Nathan B. *The Dignity of Commerce*. University of Chicago Press, 2016.

Or. Rev. Stat. Ann. § 180.545 (West 2023).

Or. Rev. Stat. § 180.545 (2005).

Parrish, Allison. "Gutenberg, dammit." .

Pew Research Center. *Social Media Fact Sheet*. April 7, 2021. .

Pilehvar, Mohammad Taher, and Jose Camacho-Collados. *Embeddings in Natural Language Processing*. 2021.

Pino, Ivana. "Credit Card Ownership and Usage Statistics." *Bankrate.com*, February 23, 2022. .

Plain Writing Act of 2010, Pub. L. No. 111-274, 124 Stat. 2861.

Plate, Norman E. "Do As I Say, Not As I Do: A Report Card on Plain Language in the United States Supreme Court." *Thomas M. Cooley Journal of Practical & Clinical Law* 13, no. 1 (2010): 79.

Plavén-Sigray, Pontus, et al. "The Readability of Scientific Texts is Decreasing Over Time." *eLife* 6 (2017).

Pozen, David E. "Transparency's Ideological Drift." *Yale Law Journal* 128, no. 1 (2018): 100.

---. "Boilerplate Today: The Rise of Modularity and the Waning of Consent." *Michigan Law Review* 104, no. 5 (2006): 1223.

Providing Accountability Through Transparency Act of 2023, Pub. L. 118-9, 137 Stat. 55.

Pub. L. No. 111-24, 123 Stat. 1734, codified in relevant part to 15 U.S.C. §§ 1601–1667(f), 1681 et seq., and 1693 et seq.

Radin, Margaret Jane. *Boilerplate: The Fine Print, Vanishing Rights, and the Rule of Law*. Princeton University Press, 2014.

Rakoff, Todd D. "Contracts of Adhesion: An Essay in Reconstruction." *Harvard Law Review* 96, no. 6 (1983): 1173.

Ramachandran, Raji, Sanju Jayachandran, and Vaishnav Das. "A Novel Method for Text Summarization and Clustering of Documents." *Proceedings of the 2022 IEEE 3rd Global Conference for Advancement in Technology*, 2022, 1.

Ramos, Juan. "Using TF-IDF to Determine Word Relevance in Document Queries." *Proceedings of the First Instructional Conference on Machine Learning*, 1999.

Readable. "Readable Pro." .

Real Estate Settlement Procedures Act, 12 U.S.C. §§ 2601–2617 (1975).

Reardon, S. F., R. A. Valentino, and K. A. Shores. "Patterns of Literacy among U.S. Students." *The Future of Children* 22, no. 2 (2012): 17. .

Restatement (Second) of Contracts § 79 (1981).

Rothwell, Jonathan. "Eradicating Illiteracy Nationally and Regionally in the United States." *Barbara Bush Foundation for Family Literacy & Gallup*, September 8, 2020. .

Rustad, Michael L. "Consumer Protection and the New Financial Products: Regulatory Reform for a Better Future." *Harvard Journal on Legislation* 54, no. 2 (2017): 543.

Rustad, Michael L. "Why a New Deal Must Address the Readability of U.S. Consumer Contracts." *Cardozo Law Review* 44 (2022): 521–31.

Rustad, Michael L., and Thomas H. Koenig. "Wolves of the World Wide Web: Reforming Social Networks' Contracting Practices." *Wake Forest Law Review* 49, no. 5 (2014): 1431.

Schmitz, Amy J. "Pizza-Box Contracts: True Tales of Consumer Contracting Culture." *Wake Forest Law Review* 45 (2010): 863–92.

Schmitz, Benedikt, and Charlotte Pavillon. "Measuring Transparency in Consumer Contracts: The Usefulness of Readability Formulas Empirically Assessed." *Journal of European Consumer & Market Law* 9, no. 3 (2020): 191–204.

Schriver, Karen. "Plain Language in the US Gains Momentum: 1940–2015." *IEEE Transactions on Professional Communication* 60 (2017): 343–83.

Schwarcz, Daniel (2011). Reevaluating standardized insurance policies. *U. Chi. L. Rev.*, *78*, 1263, 1315–17.

Schwarcz, Daniel. "Transparently Opaque: Understanding the Lack of Transparency in Insurance Consumer Protection." *UCLA Law Review* 61 (2014): 394–421.

Smith, George P., II, and William P. Lane. "Re-Evaluating the Demise of the Average, Ordinary, Reasonable Person: Unintended Consequences in the Law of Nuisance." *Catholic University Law Review* 67 (2018): 699–708.

Stark, Debra Pogrund, and Jessica M. Choplin. "A License to Deceive: Enforcing Contractual Myths Despite Consumer Psychological Realities." *New York University Journal of Law and Business* 5 (2009): 617–95.

Stark, Jack. "Should the Main Goal of Statutory Drafting Be Accuracy or Clarity?" *Statute Law Review* 15 (1994): 207–13.

Swift, Jonathan. *Gulliver's Travels*. 2021, 154. (First edition, 1726).

Touvron, Hugo, et al. *LLaMA: Open and Efficient Foundation Language Models*. Cornell University Computer Science Working Paper, February 27, 2023.

Van Boom, W. H., Desmet, P., & Van Dam, M. (2016). If it's easy to read, it's easy to claim: The effect of the readability of insurance contracts on consumer expectations and conflict behavior. J. Consumer Pol'y, 39, 187.

Watson, Amy. "Reading Habits in the U.S. – Statistics & Facts." *Statista*, January 24, 2023.

Wydick, Richard C. Plain English for Lawyers. Durham, NC: Carolina Academic Press, 2005.

White, Alan M., and Cathy Lesser Mansfield. "Literacy and Contract." *Stanford Law and Policy Review* 13 (2002): 233–36.

Wilkinson-Ryan, Tess. "A Psychological Account of Consent to Fine Print." *Iowa Law Review* 99 (2014): 1745–65

Wittgenstein, Ludwig. *Remarks on the Foundations of Mathematics*. Edited by G. H. von Wright et al., translated by G. E. M. Anscombe, revised edition, 1978, 21–28.

Zamir, Eyal. "The Inverted Hierarchy of Contract Interpretation and Supplementation." *Columbia Law Review* 97 (1997): 1710–55.

"Readability of Insurance Policies." *REGSURVEYS*, no. 0110 (2007): 42.