



# Even laypeople use legalese

Eric Martínez<sup>a,1</sup> , Francis Mollica<sup>b</sup> , and Edward Gibson<sup>c</sup> 

Edited by Timothy Wilson, University of Virginia, Charlottesville, VA; received March 18, 2024; accepted July 15, 2024

Whereas principles of communicative efficiency and legal doctrine dictate that laws be comprehensible to the common world, empirical evidence suggests legal documents are largely incomprehensible to lawyers and laypeople alike. Here, a corpus analysis ( $n = 59$  million words) first replicated and extended prior work revealing laws to contain strikingly higher rates of complex syntactic structures relative to six baseline genres of English. Next, two preregistered text generation experiments ( $n = 286$ ) tested two leading hypotheses regarding how these complex structures enter into legal documents in the first place. In line with the *magic spell hypothesis*, we found people tasked with writing official laws wrote in a more convoluted manner than when tasked with writing unofficial legal texts of equivalent conceptual complexity. Contrary to the copy-and-edit hypothesis, we did not find evidence that people editing a legal document wrote in a more convoluted manner than when writing the same document from scratch. From a cognitive perspective, these results suggest law to be a rare exception to the general tendency in human language toward communicative efficiency. In particular, these findings indicate law's complexity to be derived from its performativity, whereby low-frequency structures may be inserted to signal law's authoritative, world-state-altering nature, at the cost of increased processing demands on readers. From a law and policy perspective, these results suggest that the tension between the ubiquity and impenetrability of the law is not an inherent one, and that laws can be simplified without a loss or distortion of communicative content.

law | psycholinguistics | law and cognition | empirical legal studies | experimental jurisprudence

Since the dawn of modern civilization, humankind has codified and communicated societal norms and rules largely in the form of written laws. In order for people to understand and comply with social norms and rules, it follows that legal content must be drafted in a way such that people can ultimately understand and comply with it.

Indeed, the principle that law should provide such “fair notice” to the general public is a core tenet of modern legal doctrine, which mandates that laws provide proper warning of prohibited conduct “in language that the common world will understand,” (1, 2) to “give the person of ordinary intelligence a reasonable opportunity to know what is prohibited, so that he may act accordingly.” (3, 4).

In addition to legal doctrine, principles of communicative efficiency likewise suggest that laws should be understandable. For example, a burgeoning psycholinguistics literature has uncovered various properties of human language that appear optimized for easing the communicative burden on speakers and listeners (5–16), such as (a) syntactic dependency length minimization (17, 18), and (b) a preference for shorter words over longer words in everyday speech (19).

These principles notwithstanding, legal documents have long been observed to be notoriously difficult to understand (20–23). In particular, recent work has revealed legal documents, including both private contracts and federal legislation, to be laden with center-embedded clauses at a rate twice as high as other genres of texts, including those aimed at an educated audience (24, 25).

Moreover, legal documents containing these features have been shown to cause processing difficulty relative to legal documents without these features, even for lawyers and experienced lay readers (24, 26).

The mismatch between the ubiquity and impenetrability of legal documents has long been acknowledged not just by those tasked with reading legal documents but those tasked with promulgating them, as well (27). In the United States, policy efforts to simplify laws have been advocated for decades (28–30), with little to no success (25).

And although recent work has revealed that even lawyers prefer simplified legal documents over complex legal documents (26), it remains an open question how complex features such as center-embedded syntax make their way into legal documents in the first place.

## Significance

Why are laws so complicated? Across two preregistered experiments, we found that people tasked with writing official laws wrote in a more convoluted manner than when tasked with writing unofficial legal texts of equivalent conceptual complexity. This tendency held constant, regardless of whether people wrote the document iteratively or from scratch. These results suggest law to be a rare exception to the general tendency in human language toward communicating efficiently, and that convoluted structures may be inserted to effectively signal the authoritative nature of the law, at the cost of increased reading difficulty. These results further suggest laws can be effectively simplified without a loss or distortion of communicative content.

Author affiliations: <sup>a</sup>University of Chicago Law School, Chicago, IL 60637; <sup>b</sup>School of Psychological Sciences, University of Melbourne, Melbourne, VIC 3010, Australia; and <sup>c</sup>Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139

Author contributions: E.M., F.M., and E.G. designed research; E.M. performed research; E.M. and F.M. analyzed data; and E.M., F.M., and E.G. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

Copyright © 2024 the Author(s). Published by PNAS. This article is distributed under [Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 \(CC BY-NC-ND\)](https://creativecommons.org/licenses/by-nc-nd/4.0/).

<sup>1</sup>To whom correspondence may be addressed. Email: [ericmart@mit.edu](mailto:ericmart@mit.edu).

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2405564121/-/DCSupplemental>.

Published August 19, 2024.

To answer this question, we conducted two well-powered preregistered experiments testing two leading hypotheses for why lawyers write the way that they do, including: (a) the magic spell hypothesis, according to which lawyers and lawmakers write in a convoluted manner in order to lend legal documents a ritualistic, spell-like element; and (b) the copy-and-edit hypothesis, according to which conditions and specifications are often considered only after the creation of an initial draft and are more easily embedded into the center of existing sentences as opposed to being written-out into separate sentences.

In line with the magic spell hypothesis (22, 25), we found that people tasked with writing laws wrote in a more convoluted manner (i.e. more center-embedded syntax) than when tasked with writing control texts of plausibly equivalent conceptual complexity. Contrary to the copy-and-edit hypothesis, we did not find evidence that people editing a legal document wrote in a more convoluted manner than when writing the document from scratch.

These findings suggest that lawyers and lawmakers write in a complex manner in order to confer legal documents a ritualistic, spell-like element, presenting broad-ranging implications for law, policy, and cognitive science.

### Law's Syntactic Complexity

Perhaps the most distinctive feature of legalese is center-embedded syntax, in which clausal content is embedded within the center of another clause as opposed to being edge-embedded or written as a separate sentence.

Consider the following example from a Massachusetts Drunk Driving Law:

*“Whoever, upon any way or in any place to which the public has a right of access, or upon any way or in any place to which members of the public have access as invitees or licensees, operates a motor vehicle with a percentage, by weight, of alcohol in their blood of eight one-hundredths or greater, or while under the influence of intoxicating liquor, or of marijuana, narcotic drugs, depressants, or stimulant substances, all as defined in section one of chapter ninety-four C, or while under the influence from smelling or inhaling the fumes of any substance having the property of releasing toxic vapors as defined in section 18 of chapter 270 shall be punished by a fine of not less than five hundred nor more than five thousand dollars or by imprisonment*

*for not more than two and one-half years, or both such fine and imprisonment.”* (31)

The subject of this provision (“Whoever”) and the main verb phrase (“shall be punished”) are separated by over 100 words of clausal material, resulting in a center-embedded structure that is unusually difficult to process (18, 32, 33).

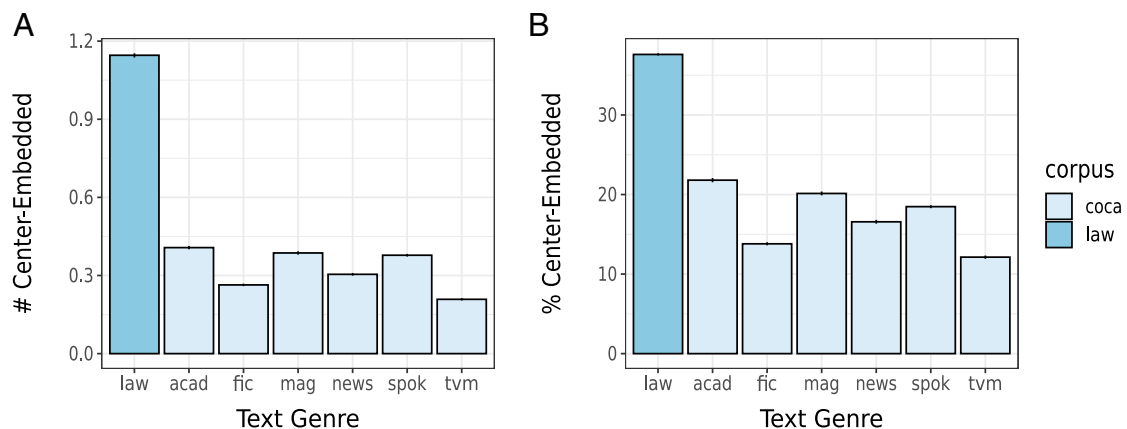
Prior work has indicated that this example is by no means unique, as legal documents have been found to contain strikingly higher rates of center-embedded syntax relative to other genres of English, including those aimed at an educated audience (24, 25).

For robustness purposes, here we first sought to replicate and extend these results using a more direct method of detecting center-embedded syntax compared to prior work (*Materials and Methods*), in which we used state-of-the-art natural language processing tools to detect the number of center-embedded verbs in a sentence in (a) the United States Code (34) and (b) six baseline genres in the Corpus of Contemporary American English (35): academic texts, fiction, newspaper articles, magazine articles, spoken transcripts, and TV/Movie scripts.

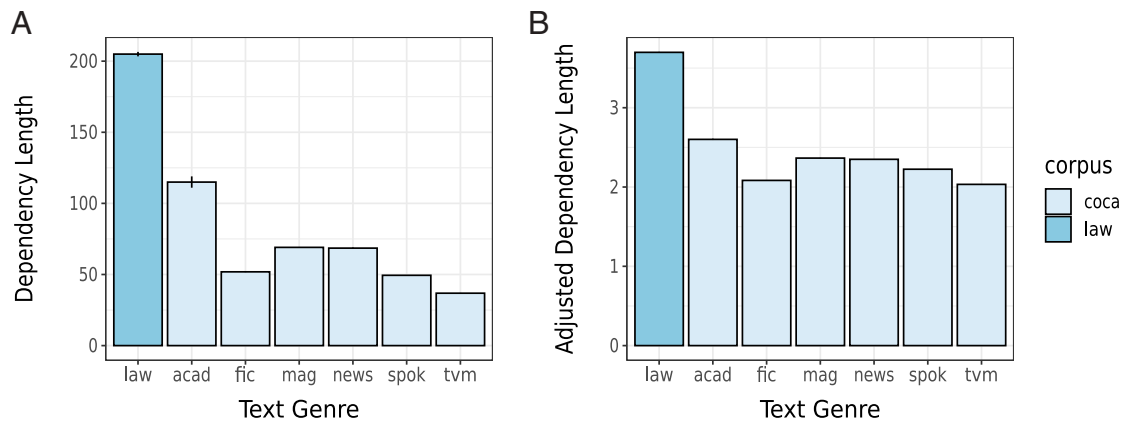
Results are visualized in Fig. 1. Consistent with prior work, laws contained several times more center-embedded clauses than any of the baseline genres of English. When looking at the percentage of sentences with center-embedded clauses, laws likewise contained strikingly higher rates than any other genre.

In addition, prior analyses have also indicated that center-embedded syntax disproportionately contributes to the higher difficulty in recalling legalese versus plain-English compared to other markers of legalese, such as passive voice and nonstandard capitalization (24). The increased processing difficulty associated with center-embedded syntax in legal texts and nonlegal texts has been hypothesized to be associated with increased demands on working memory capacity resulting from long-distance syntactic dependencies (24, 36).

As explained in ref. 17, the difficulty of long dependencies follows naturally from leading models of human language processing. Dependency length corresponds to the length of time in which a word must be held in memory in a left-corner parser or generator (37–39). Given that storing items in memory may be difficult or error-prone, short dependencies by hypothesis would be easier and more efficient to produce and parse. In support of this idea, comprehension and production difficulty have been observed at the sites of long dependencies (36, 40).



**Fig. 1.** Number of center-embedded clauses per sentence (A) and percentage of sentences with center-embedded syntax (B) in laws compared to six baseline genres of written and spoken English: academic texts, fiction, magazine articles, newspaper articles, and TV/Movies. Laws were taken from the 2021 edition of the United States Code, the official compilation of all federal laws currently in force. Baseline genres were taken from the most recent year (2019) of the Corpus of Contemporary American English. Error bars represent 95% bootstrapped CIs.



**Fig. 2.** Dependency length (A) and adjusted dependency length (B) in laws versus academic texts, fiction texts, newspaper articles, magazine articles, spoken transcripts, and TV/Movie scripts.

Moreover, previous literature has shown not only that long-distance dependencies are difficult to produce and comprehend relative to shorter dependencies, but also that they are widely dispreferred in language usage relative to shorter dependencies. For example, (17) found across 37 natural languages that utterances contained significantly shorter dependencies relative to several conservative baselines.

If legal documents had longer syntactic dependencies than baseline, it would suggest that legal text imposes greater memory/processing demands on readers than other domains of language, in turn advancing our understanding of both why legal documents are difficult to process as well as the domains and degree to which language is optimized for communicative efficiency.

However, it remains an open question to what extent legal texts have longer syntactic dependencies relative to baseline texts.

To answer this question, we also compared the syntactic dependency length in our legal versus nonlegal corpora. As with center-embedded syntax, and consistent with the predictions of the theoretical literature, laws contained strikingly longer dependencies than any of the other baseline genres. Results are visualized in Fig. 2.

## Hypotheses

Having replicated and extended prior work demonstrating the prevalence of complex syntactic structures in legal texts, we next turned to testing two nonmutually exclusive hypotheses proposed in the previous literature for how such features enter into legal documents in the first place. Below we briefly present each of these hypotheses in turn, as well as the associated predictions of these hypotheses that we preregistered for our experiments.

**Magic Spell Hypothesis.** Some have posited that lawyers and lawmakers write in a convoluted manner in order to lend legal documents a ritualistic, spell-like element (22, 25). These ritualistic types of language are often referred to as performative utterances (41), which unlike descriptive utterances, not only describe the state of the world but change the state of the world they are describing.

In order to effectively convey performativity, such utterances have been attested to contain distinctive, low-frequency structures, as in the case of magic spells, which are characterized by such features as rhyming [e.g. “Double, double toil and trouble; Fire burn, and cauldron bubble”: (42)] and foreign-sounding jargon [“wingardium leviosa” (43)]. Indeed, in a pilot experiment, we found that participants tasked with writing

a magic spell rhymed in 58.8% of sentences, as compared to 1.8% of sentences when tasked with writing a mere recollection of a fantastical event involving a magic spell (*SI Appendix*).

Given that legal documents, like spells and other performative utterances: (a) have been shown to possess certain distinctive low-frequency structures (such as center-embedded syntax), at several times the rate of standard texts (24, 25), and (b) are meant not only to describe the state of the world but also change the state of the world (by establishing, eliminating, and/or modifying legally binding social rules), one might similarly hypothesize that such low-frequency structures are inserted so as to signal a legal document’s authoritative nature.

If this hypothesis were true, one would predict that people tasked with writing an official legal document would write in a more convoluted manner (including more center-embedded syntax) than when writing a nonperformative law-related document of equivalent conceptual complexity.

**Copy-and-Edit Hypothesis.** Recent work has speculated that convoluted legal language may be a result of an iterative drafting process, in which conditions and specifications are often thought of only after the creation of an initial draft or template and are more easily embedded into the center of existing sentences as opposed to being written-out into separate sentences (25).

Although this hypothesis may not fully account for all aspects of convoluted language in legal texts (such as low-frequency jargon), this would explain why the prevalence of structures such as center-embedded syntax is so much higher in legal documents than other genres where the drafting process is less path-dependent and drawn-out (24, 25), particularly given the observed reliance of lawyers and lawmakers on templates and “boilerplate provisions” in the drafting of legal documents (44, 45).

If this hypothesis were true, one would predict that people tasked with editing a legal document will write in a more convoluted manner (including more center-embedded clauses) than when tasked with writing a legal document of equivalent semantic content from scratch.

## Results

**Experiment 1.** In Experiment 1, we evaluated both the magic spell Hypothesis and copy-and-edit hypothesis.\* To evaluate the

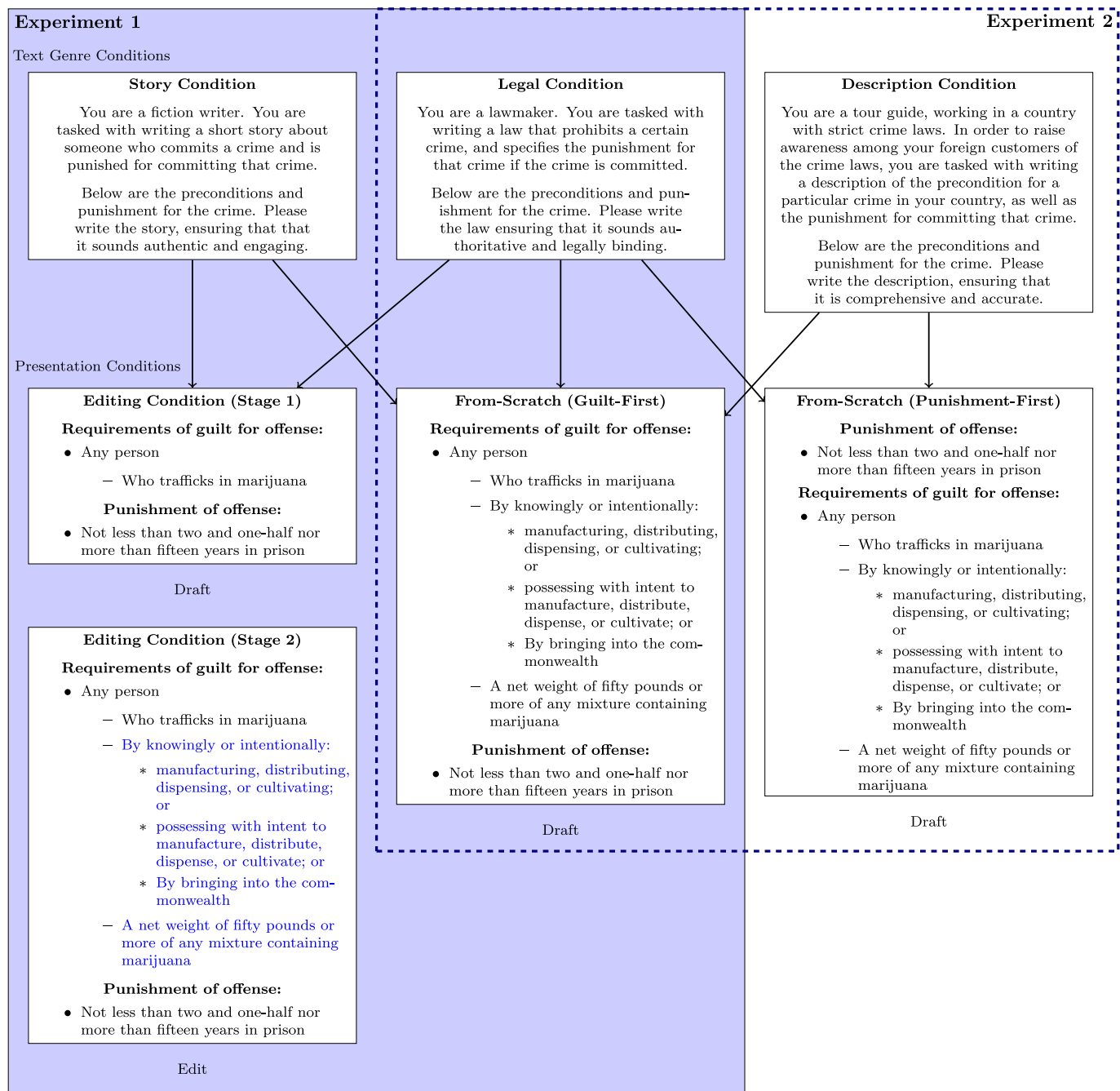
\* All data and code for this paper can be viewed at the following Open Science Framework (OSF) repository link: [link](#). The preregistration for Experiment 1 can be found at the following link: [link](#).

predictions of these hypotheses, we conducted a preregistered within-subject experiment in which we asked participants ( $n = 204$ ) to write both (a) legal provisions prohibiting crimes; and (b) stories describing someone committing crimes. See Fig. 3 for overview of experimental design.

In half of the trials (from-scratch condition), participants were initially given all of the details of the crime and were tasked with writing their law or story all at once. In the other half of trials (editing condition), participants were first given details of a paired-down version of the crime and were

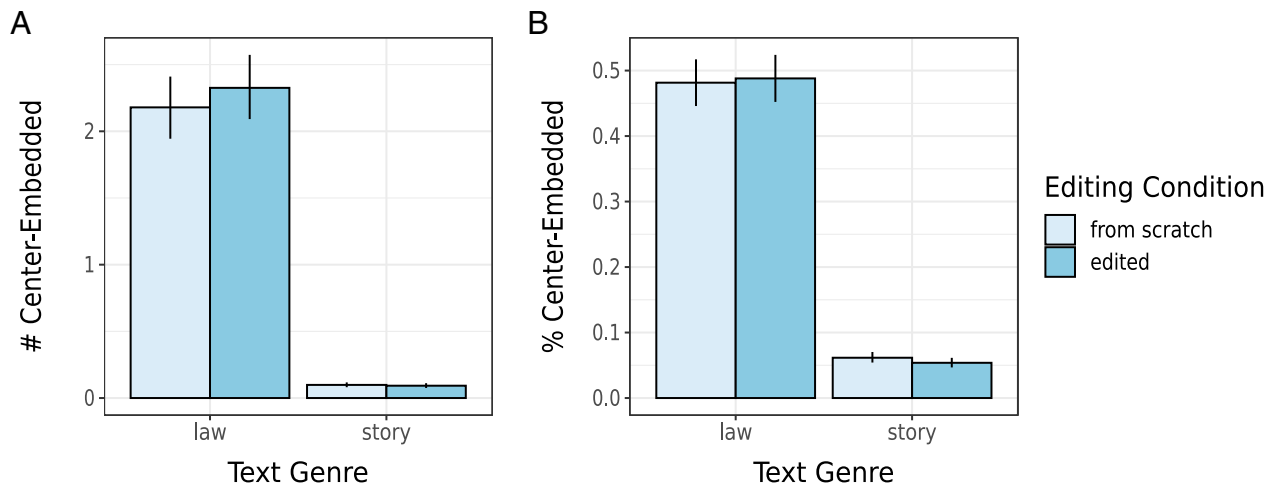
asked to write their law or story based on that version. After completing their draft, participants in these trials were then presented with additional details of the crime and were asked to revise their draft to incorporate these additional details. Conditions were manipulated within-participant and across crimes (*SI Appendix*).

Results are visualized in Fig. 4. In line with the predictions of the magic spell hypothesis, participants' responses contained a higher percentage of sentences with center-embedded syntax in the law condition (48.1%; 95% CI: 46.0 to 51.1) compared to



**Fig. 3.** Schema of experimental design. On each trial, a participant would be given instructions asking them to either write a law prohibiting a crime (*legal condition*), a story describing the commission of a crime (Experiment 1 *control condition*) or a description of a crime (Experiment 2 *control condition*). Participants would then be given a set of legal propositions to be included in the law or story/description. In the *from-scratch condition* (Experiment 1 and 2), these propositions would be presented all at once followed by a response box. In the *editing condition* (Experiment 1), a paired-down set of propositions would be presented followed by a response box. After a response is made, the full set of propositions are revealed and participants are instructed to update their response.





**Fig. 4.** Number of center-embedded clauses per sentence (A) and percentage of sentences with center-embedded syntax (B) in criminal laws versus crime stories. Error bars represent 95% bootstrapped CIs.

responses in the story condition (5.8%; 95% CI: 5.2 to 6.2). The difference was striking [OR: 8.3], and significant ( $\beta = 2.859$ , SE = 0.113,  $P < 0.0001$ ), and held true when looking at the number of center-embedded clauses per sentence ( $\beta = 3.126$ , SE = 0.204,  $P < 0.0001$ ) as opposed to just the percentage of sentences with center-embedded clauses (see *SI Appendix* for full details of model structure).

Contrary to the predictions of the copy-and-edit hypothesis, participants in the editing condition were not significantly more likely to center-embed than in the from-scratch condition ( $P = 0.262$ ), nor was there an interaction between genre and editing manipulations ( $P = 0.244$ ). This was true both when looking at the number of center-embedded clauses per sentence and when examining the percentage of sentences with center-embedded clauses ( $P = 0.755$  for editing manipulation;  $P = 0.165$  for interaction between editing and genre manipulations).

**Experiment 2.** To further test the robustness of the magic spell hypothesis, we conducted a second within-subject experiment in which we asked participants ( $n = 82$ ) to write either (a) an official law prohibiting a crime (law condition); or (b) an unofficial description of a law prohibiting a crime (description condition), with the latter being a plausibly tighter control than a story for a text of similar conceptual complexity as a law.<sup>†</sup>

This experiment also contained a second manipulation whose purpose was to control for the possibility that people might be more likely to center-embed depending on the order in which the details of the crime in question were presented to the participant.

In particular, one might conceivably believe that the presentation of propositions in Experiment I (requirements of guilt, then punishment) would have primed participants to center-embed the guilt-related propositions relative to an alternative presentation of the propositions.

To that end, this second manipulation consisted of a guilt-first condition and a punishment-first condition. In the guilt-first condition, the instructions described the requirements of guilt for the prohibited crime, followed by the punishment for the crime. In the punishment-first condition, the instructions described the punishment for the crime followed by the requirements of guilt for the crime.

<sup>†</sup>The preregistration for Experiment II can be viewed at the following link: [link](#).

Results are visualized in Fig. 5. As in Experiment 1, in line with the magic spell hypothesis, participants were more likely to produce sentences containing center-embedded clauses in the law condition (54.6%; 95% CI: 50.3 to 59.1) than in the control condition (25.7%; 95% CI: 22.5 to 28.9). The difference was striking [OR: 2.1], and was significant both when looking at the number of center-embedded clauses per sentence ( $\beta = 1.391$ , SE = 0.184,  $P < 0.0001$ ) as well as the percentage of sentences with center-embedded clauses ( $\beta = 1.552$ , SE = 0.227,  $P < 0.0001$ ).

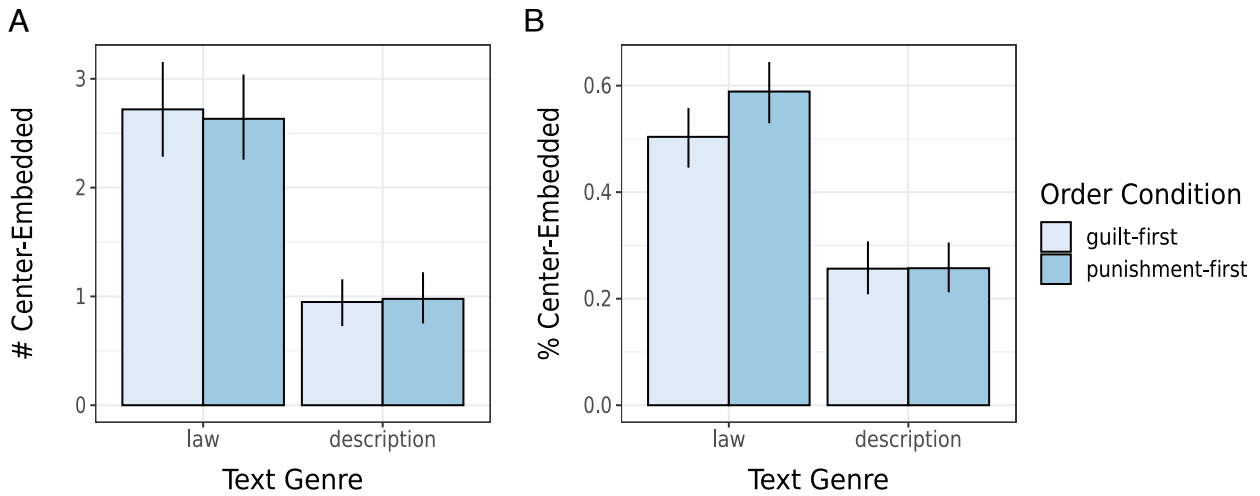
The results of the ordering manipulation were also consistent with the magic spell hypothesis, as participants were not significantly more likely to produce sentences with center-embedded syntax in the guilt-first condition ( $P = 0.613$ ) than in the punishment-first condition, nor was there an interaction between genre and ordering manipulation ( $P = 0.414$ ). Converging results were found when analyzing the number of center-embedded clauses per sentence ( $P = 0.362$  for ordering manipulation;  $P = 0.274$  for interaction between ordering and genre manipulations).

To further test the robustness of the magic spell hypothesis and account for the possibility that participants' responses in the two conditions were not matched for conceptual complexity, we conducted additional exploratory analyses where (a) laws were coded into legally relevant propositions as in ref. 24, (b) responses were filtered if they did not include more than 80% of the propositions in the instructions, and (c) conceptual complexity (operationalized as proportion of propositions included in a participant's response) was included as a fixed-effect predictor in our regression models.

The results of these analyses were consistent with those reported in the main text; genre remained a significant and strong predictor of participants' likelihood to center-embed. That is, participants continued to center-embed at a higher rate in laws versus unofficial descriptions of laws, even when removing responses that were less conceptually complex and when explicitly accounting for conceptual complexity as a predictor variable. These analyses are reported in full in *SI Appendix*.

## Discussion

This paper has empirically investigated the long-puzzling question of why laws are written in a complex manner, testing



**Fig. 5.** Number of center-embedded clauses per sentence (A) and percentage of sentences with center-embedded syntax (B) in participant-drafted laws versus unofficial descriptions of laws. Error bars represent 95% bootstrapped CIs.

two leading hypotheses across two well-powered, preregistered experiments.

In line with the magic spell hypothesis, we found that people tasked with drafting laws wrote in a more convoluted manner than when tasked with drafting various control texts of plausibly equivalent conceptual complexity. Contrary to the copy-and-edit hypothesis, we did not find evidence that people editing a legal document wrote in a more convoluted manner than when writing the document from scratch.

These lines of evidence were robust to various control attempts, including (a) comparing laws to different genres (stories and descriptions of laws) to serve as control texts; and (b) manipulating the order in which propositions of a given law were presented (requirements of guilt first versus punishment first).

Answering this question is relevant to advancing longstanding questions of both cognitive science and legal doctrine/public policy.

On the cognitive science side, as documented above, there is a burgeoning psycholinguistics literature documenting the various domains in which communicative efficiency shapes human language (5–16). Given that law stands as an attested exception to this observed efficiency, uncovering the cognitive factors giving rise to the processing difficulties of legal documents can help inform the degree and domains in which human language is optimized for communicative efficiency, as well as the factors giving rise to said (in)efficiency.

In particular, these results suggest law to be a type of performative utterance (41), meant not just to communicate states of the world but to explicitly alter the state of the world. In such instances, distinctive low-frequency structures may be inserted in order to effectively signal the performative nature of the utterance, which in turn might increase processing demands on readers. In the case of other types of performative language, such as “actual” magic spells, such structures may include rhyming or foreign-sounding terminology. In the case of laws, this deviation may come largely in the form of altering the syntactic structure of the clausal material from right-branching to center-embedded, creating as a by-product an overload on a reader’s working memory capacity.

On the law and policy side, these results add to an emerging body of literature demonstrating that the language of legal documents can be simplified without a loss or distortion of legal

content (24–26), which might provide a source of optimism to efforts to simplify legal documents [which have been advocated for for decades (46), to no avail (25)]. These findings also shed insight into debates related to the aforementioned legal doctrines that expressly assert or implicitly assume that laws be understandable to the public at large. Jurists have long acknowledged the tension between the doctrinal mandate that laws be understandable to the common person and the observation that laws are not understandable to the common person (1–4). Whereas recent proposals to resolve this tension have taken for granted the necessity of law’s complexity and have called for scaling back the mandate that laws be accessible to the common person (47), our results suggest such compromises may not be necessary. Instead, our results indicate that lawmakers can faithfully comply with this mandate while simultaneously preserving the desired level of conceptual complexity.

**Ethics Approval.** Ethics Approval. Both experiments were approved by Massachusetts Institute of Technology’s Committee on the Use of Humans as Experimental Subjects (COUHES), protocol number: 2107000425. Prior to completing each experiment, participants were shown a consent form, which provided further details about the experiment, including risks, as well as information about COUHES approval.

## Materials and Methods

### Corpus Analysis.

**Materials.** Our primary materials consisted of two corpora. Our legal corpora consisted of the 2021 edition of the United States Code (34), the official compilation of all federal legislation currently in force.

Our baseline corpora consisted of academic texts, fiction texts, newspaper articles, magazine articles, spoken transcripts, and TV/Movie scripts from the Corpus of Contemporary American English (35). In order to best match the legal corpora, we used only texts from the most recent year of the corpus (2019).

**Procedure.** To calculate the number of center-embedded clauses in each sentence, we first (a) tokenized each corpus into sentences; and (b) got a syntactic parse of the sentence using the Stanza package from the Stanford NLP group (48). Following refs. 24 and 25, we then filtered out sentences that i) contained fewer than 10 alphabetic words; ii) did not end in a punctuation mark; or iii) contained 3 or more punctuation characters in a row.

For each sentence, we then calculated the number of center-embedded verbs (operationalized as the number of main verbs between a noun and its root).

For validation purposes, we hand-coded a random sample of 300 sentences for the presence of center-embedded clauses. This revealed the parser to be 92.3% accurate in detecting whether a sentence contained a center-embedded clause (95% CI: 89.3 to 95.5).

For the dependency analysis, for each word, we calculated the distance between that word and its head word (defined as the difference in ordinal position/index between the word and its head word).

For each sentence, we then calculated (a) the total distance across words; and (b) the total distance across words divided by the total number of words.

### Experiment 1.

**Materials.** The primary materials consisted of eight items, with each item consisting of sets of instructions to write a passage relating to (respectively) the commission of a legally prohibited criminal offense (i.e. a crime), such as arson, bribery, or drunk driving. Each item consisted of 4 conditions (2 manipulations with 2 conditions each). The first manipulation was genre, which consisted of a legal condition and a story condition. In the legal condition, the materials consisted of instructions asking participants to write a law prohibiting a crime. In the story condition, participants were asked to write a story involving someone committing a crime. Both conditions had an associated cover story explaining the motivation behind the task. In the legal condition, participants were told that they were a "lawmaker" who was "tasked with writing a law that prohibits a certain crime, and specifies the punishment for that crime if the crime is committed." In the story condition, participants were told that they were a "fiction writer" who was "tasked with writing a story about someone who commits a crime and is punished for committing the crime."

The second manipulation was sequencing, whose conditions consisted of a from-scratch condition and an editing condition. In the from-scratch condition, the details and propositions of the crime were presented all at once.

In the editing condition, in contrast, the propositions were presented in two stages. In Stage 1 (Fig. 3), the version of the crime included within the instructions was paired-down and did not contain all of the propositions. In Stage 2, the version of the crime included all of the propositions, and the instructions directed participants to edit their text so as to include all of the additional instructions (see blue text of Fig. 3).

**Participants and procedure.** Participants were recruited via the online platform Prolific. To compute sample size, we conducted a power analysis, which determined the sample that would give us an 80% chance to detect an effect size that was at least half as large as the effect of the interaction between genre + sequencing obtained in a pilot experiment (this was smallest effect of any predictor variable in our pilot experiment). Based on this power analysis, and assuming a 10% exclusion rate, we recruited 220 participants.

Participants were eligible if they resided in the United States, were 18 years or older, and native speakers of English. Each participant completed 8 trials of the same series of tasks.

On a given trial, participants would be presented with materials in one of the four conditions and asked to write either a law or story in accordance with the material's instructions. As noted above, when in the from-scratch condition, participants were asked to draft their text all at once, whereas in the editing condition, participants were first asked to write an initial draft based on a paired-down version of the crime described, and then subsequently presented with the full version of the crime and asked to edit their draft to incorporate the additional details associated with that version. Across the 8 trials, each participant was presented with 2 items in each of the 4 conditions, never seeing the same item more than once.

Prior to each trial, participants were given a comprehension check question where they were (a) told which of the two genres they would be asked to write (a law or a story), and (b) asked to confirm which of the two genres they would be asked to write. Participants were not allowed to proceed to the trial until answering the comprehension check correctly.

Prior to completing the first trial, participants were asked to promise that they would not use a language model [such as generative pre-trained transformer (GPT)] to complete the task. After completing the last trial, they were prompted with a similar message asking to promise that they did not use a language model (such as GPT) to complete the task.

Participants were retained in the analysis if they completed all trials and were determined not to use a language model in their responses. Based on

these criteria, 204 of the 220 recruited participants were retained in the final analysis.

**Analysis plan.** To evaluate participant responses, responses were separated into sentences using an automatic parser—in particular, the tokenizers package in R. The tokenized sentences were spot-checked by a human and corrected for errors. After tokenization, sentences were hand-coded for center-embedded syntax, both in terms of (a) the degree of center-embedded syntax (defined as the number of center-embedded verbs); and (b) the binary presence of center-embedded syntax (i.e. were any verbs in the sentence center-embedded).

Following our preregistration, we then analyzed the effect of our two manipulations on the prevalence of center-embedded syntax by conducting two separate regressions for each of the two operationalizations of center-embedded syntax, including (a) a mixed-effects binary logistic regression with the binary presence of center-embedded syntax (in a given sentence) as the outcome variable; and (b) a mixed-effects Poisson regression with degree of center-embedded syntax as the outcome variable. Both regressions featured (a) genre, sequencing condition and their interaction as fixed-effects; and (b) genre, sequencing condition, item, and participant as random intercepts.

Results did not qualitatively change for either regression. We report both in the text.

### Experiment 2.

**Materials.** Similar to Experiment 1, the primary materials of Experiment 2 consisted of eight items, each of which consisted of 4 conditions (2 manipulations with 2 conditions each). The first manipulation was genre, which consisted of a law condition and a description condition. The law condition was identical to the law condition in Experiment 1, and consisted of instructions asking participants to write a law prohibiting a crime. In the description condition, participants were asked to write an unofficial description of a law prohibiting a crime.

As in Experiment 1, both conditions had an associated cover story explaining the motivation behind the task. As in Experiment 1, participants in the law condition were told that they were a "lawmaker" who was "tasked with writing a law that prohibits a certain crime, and specifies the punishment for that crime if the crime is committed." In the description condition, participants were told that they were a "tour guide" working in a country with strict crime laws. In order to raise awareness among foreign customers of the crime laws, they were "tasked with writing a description of the preconditions for a particular crime in your country, as well as the punishment for committing that crime."

The second manipulation was ordering, the purpose of which was to account for the possibility that people might be more likely to center-embed depending on the order in which the details of the crime in question were presented to the participant.

In particular, one might conceivably believe that the presentation of propositions in Experiment 1 (requirements of guilt, then punishment) would have primed participants to center-embed the guilt-related propositions relative to an alternative presentation of the propositions.

To that end, the ordering manipulation consisted of a guilt-first condition and a punishment-first condition. In the guilt-first condition, the details of the crime in question were presented such that the requirements of guilt for the offense were presented first, followed by the punishment of the offense. In the punishment-first condition, the ordering was reversed, such that the punishment of the offense was presented first, followed by the requirements of guilt.

Unlike Experiment 1, there was no sequencing manipulation—across all conditions, the materials asked participants to write their law or description all at once from scratch instead of in stages.

**Participants and procedure.** Participants ( $n = 82$ ) were recruited via the online platform Prolific. This sample size was based on a power analysis, which determined the number of participants that would give us an approximately 80% chance to detect an effect size that was at least 1/5 as large as the effect of genre obtained in Experiment 1. Participants were eligible if they resided in the United States, were 18 years or older, and native speakers of English. Each participant completed 8 trials of the same series of tasks.

On a given trial, participants were presented with materials in one of the four conditions, and asked to write a text of the appropriate genre. Across the 8 trials, each participant was presented with 2 items in each of the 4 conditions, never seeing the same item more than once. As in Experiment 1, participants were given a comprehension check prior to each trial, were asked before and after

the experiment to promise to not use/have used a language model to generate their responses, and were retained according to the same exclusion criteria. All 82 participants passed the exclusion criteria and were thus retained in the final analysis.

**Analysis plan.** Responses were tokenized and coded for center-embedded syntax following the same procedure as in Experiment 1. As in Experiment 1, we analyzed the effect of our two manipulations on the prevalence of center-embedded syntax by conducting two separate regressions for each of the two operationalizations of center-embedded syntax, including (a) a mixed-effects binary logistic regression with the binary presence of center-embedded syntax (in a given sentence) as the outcome variable; and (b) a mixed-effects Poisson regression with degree of center-embedded syntax as the outcome variable. Both regressions featured (a) genre, ordering, and their interaction as fixed-effects; and (b) genre, ordering, item, and participant as random intercepts. Results did not qualitatively change for either regression. We therefore report both in the text.

1. McBoyle vs. United States, 283 U.S. 552. 1931.
2. Moskal vs. United States, 498 U.S. 89-964. 1990.
3. Grayned vs. City of Rockford, 408 U.S. 104. 1972.
4. Hoffman Estates vs. Flipside, Hoffman Estates, Inc., 455 U.S. 80-1681. 1982.
5. E. Gibson *et al.*, How efficiency shapes human language. *Trends Cogn. Sci.* **23**, 389–407 (2019).
6. S. T. Piantadosi, H. Tily, E. Gibson, Word lengths are optimized for efficient communication. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 3526–3529 (2011).
7. S. T. Piantadosi, H. Tily, E. Gibson, The communicative function of ambiguity in language. *Cognition* **122**, 280–291 (2012).
8. S. T. Piantadosi, Zipf's word frequency law in natural language: A critical review and future directions. *Psychon. Bull. Rev.* **21**, 1112–1130 (2014).
9. K. Mahowald, E. Fedorenko, S. T. Piantadosi, E. Gibson, Info/information theory: Speakers choose shorter words in predictive contexts. *Cognition* **126**, 313–318 (2013).
10. K. Mahowald, I. Dautriche, E. Gibson, S. T. Piantadosi, Word forms are structured for efficient use. *Cogn. Sci.* **42**, 3116–3134 (2018).
11. R. Futrell, E. Gibson, L2 processing as noisy channel language comprehension. *Biling. Lang. Cogn.* **20**, 683–684 (2017).
12. R. Futrell and R. Levy, "Noisy-context surprisal as a human sentence processing cost model" in *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers* (Association for Computational Linguistics (ACL), 2017), vol. 1, pp. 688–698.
13. R. Ryskin, R. Futrell, S. Kiran, E. Gibson, Comprehenders model the nature of noise in the environment. *Cognition* **181**, 141–150 (2018).
14. Y. Zhang, R. Ryskin, E. Gibson, A noisy-channel approach to depth-charge illusions. *Cognition* **232**, 105346 (2023).
15. E. Gibson *et al.*, A noisy-channel account of crosslinguistic word-order variation. *Psychol. Sci.* **24**, 1079–1088 (2013).
16. E. Gibson, L. Bergen, S. T. Piantadosi, Rational integration of noisy evidence and prior semantic expectations in sentence interpretation. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 8051–8056 (2013).
17. R. Futrell, K. Mahowald, E. Gibson, Large-scale evidence of dependency length minimization in 37 languages. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 10336–10341 (2015).
18. E. Gibson, N. J. Pearlmutter, Constraints on sentence comprehension. *Trends Cogn. Sci.* **2**, 262–268 (1998).
19. G. K. Zipf, *Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology* (Addison-Wesley Press, 1949).
20. P. M. Tiersma, Reforming the language of jury instructions. *Hofstra Law Rev.* **22**, 37 (1993).
21. P. Tiersma, Some myths about legal language. *J. Law Cult. Humanit. Forthcom. Loyola-LA Leg. Stud. Pap.* **2**, 2005–2026 (2005).
22. P. Tiersma, "The nature of legal language" in *ALLA Applied Linguistics Series*, J. Gibbons, M. T. Turell, Eds. (John Benjamins Publishing Company, 2008), vol. 5, pp. 7–25.
23. M. E. J. Masson, M. A. Waldron, Comprehension of legal contracts by non-experts: Effectiveness of plain language redrafting. *Appl. Cogn. Psychol.* **8**, 67–85 (1994).
24. E. Martínez, F. Mollica, E. Gibson, Poor writing, not specialized concepts, drives processing difficulty in legal language. *Cognition* **224**, 105070 (2022).
25. E. Martínez, F. Mollica, E. Gibson, So much for plain language: An analysis of the accessibility of united states federal laws over time. *J. Exp. Psychol. Gen.* **153**, 1153–1164 (2024).
26. E. Martínez, F. Mollica, E. Gibson, Even lawyers don't like legalese. *Proc. Natl. Acad. Sci. U.S.A.* **120**, 1–7 (2023).
27. M. Adler, "The plain language movement" in *The Oxford Handbook of Language and Law*, P. M. Tiersma, L. M. Solan, Eds. (Oxford University Press, 2012).
28. Plain Language Action Information Network. Federal plain language guidelines, 2011.
29. Exec. Order No. 13648. 44 fr 69609, 1979.
30. Plain Writing Act of 2010. "5 U.S.C. Section 301. 2010".
31. Massachusetts General Laws Chapter 90, Section 24.
32. G. A. Miller, N. Chomsky, "Finitary models of language users" in *Handbook of Mathematical Psychology*, R. Luce, R. R. Bush, E. E. Galante, Eds. (John Wiley & Sons, 1963), pp. 2–419.
33. N. Chomsky, On certain formal properties of grammars. *Inf. Control* **2**, 137–167 (1959).
34. United States Code. U.S. Government Publishing Office, Washington, D.C., 2021. URL <https://www.govinfo.gov/app/collection/uscode/>2021.
35. M. Davies, The 385+ million word corpus of contemporary American English (1990–2008+): Design, architecture, and linguistic insights *Int. J. Corpus Linguist.* **14**, 159–190 (2009).
36. E. Gibson, Linguistic complexity: Locality of syntactic dependencies. *Cognition* **68**, 1–76 (1998).
37. P. Resnik, "Left-corner parsing and psychological plausibility" in *COLING 1992 Volume 1: The 14th International Conference on Computational Linguistics* (Association for Computational Linguistics (ACL), 1992).
38. E. A. F. Gibson, *A Computational Theory of Human Linguistic Processing: Memory Limitations and Processing Breakdown* (Carnegie Mellon University, 1991).
39. S. P. Abney, M. Johnson, Memory requirements and local ambiguities of parsing strategies. *J. Psycholinguist. Res.* **20**, 233–250 (1991).
40. D. Grodner, E. Gibson, Consequences of the serial nature of linguistic input for sentential complexity. *Cogn. Sci.* **29**, 261–290 (2005).
41. J. L. Austin, "Performative utterances" in *Philosophical Papers*, J. O. Urmson, G. J. Warnock, Eds. (Clarendon Press, 1961).
42. W. Shakespeare, *Macbeth* (E-Kitap Projesi & Cheapest Books, 2024).
43. J. K. Rowling, *Harry Potter and the Philosopher's Stone* (Bloomsbury Publishing, 2015), vol. 1.
44. J. Nyarko, Stickiness and incomplete contracts. *U. Chi. L. Rev.* **88**, 1 (2021).
45. I. V. Robert Anderson, Path dependence, information, and contracting in business law and economics. *Wis. L. Rev.* **2020**, 553 (2020).
46. M. A. Blasie, The rise of plain language laws. *U. Miami L. Rev.* **76**, 447 (2021).
47. J. Cross, The fair notice fiction. *Ala. Law Rev.* **75**, 83 (2023).
48. P. Qi, Y. Zhang, Y. Zhang, J. Bolton, C. D. Manning, Stanza: A python natural language processing toolkit for many human languages. arXiv [Preprint] (2020). <http://arxiv.org/abs/2003.07082> (Accessed 28 July 2024).
49. E. Martínez, F. Mollica, E. Gibson, Data from "Even Laypeople Use Legalese." OSF. [https://osf.io/p64h2/?view\\_only=95f1031413b14825aaef3b8b2d3e9617](https://osf.io/p64h2/?view_only=95f1031413b14825aaef3b8b2d3e9617). Accessed 28 July 2024.
50. E. Martínez, F. Mollica, E. Gibson, Magic Spell Experiment 1. OSF. [https://osf.io/6g9va/?view\\_only=8b53849820d34dc186df57918c0c18bb](https://osf.io/6g9va/?view_only=8b53849820d34dc186df57918c0c18bb). Accessed 28 July 2024.
51. E. Martínez, F. Mollica, E. Gibson, Magic Spell Experiment 2. OSF. [https://osf.io/4tfqx/?view\\_only=82a899c7a105465382c7b0c7439a0f19](https://osf.io/4tfqx/?view_only=82a899c7a105465382c7b0c7439a0f19). Accessed 28 July 2024.

**Data, Materials, and Software Availability.** Anonymized; pre-registration for Experiment 1; pre-registration for Experiment 2 data have been deposited in OSF ([https://osf.io/p64h2/?view\\_only=95f1031413b14825aaef3b8b2d3e9617](https://osf.io/p64h2/?view_only=95f1031413b14825aaef3b8b2d3e9617), [https://osf.io/6g9va/?view\\_only=8b53849820d34dc186df57918c0c18bb](https://osf.io/6g9va/?view_only=8b53849820d34dc186df57918c0c18bb), and [https://osf.io/4tfqx/?view\\_only=82a899c7a105465382c7b0c7439a0f19](https://osf.io/4tfqx/?view_only=82a899c7a105465382c7b0c7439a0f19)) (49–51).

**ACKNOWLEDGMENTS.** We thank Goeun Kim for help in drafting the materials. We thank Rebecca Saxe for the inspiration for testing and naming the magic spell hypothesis. We thank Nancy Kanwisher, Roger Levy, Ev Fedorenko, Laura Schulz, and Claire Hill for helpful feedback and discussions regarding the hypotheses and experimental design. We thank audiences at Tedlab and Evlab, the Massachusetts Institute of Technology Brain and Cognitive Sciences "Cognitive Lunch" talk series, and the Western, Educated, Industrialized, Rich, Democratic conference in Minneapolis, MN October 2023.