Quantifying Vision through Language Demonstrates that Visionary Ideas Come from the Periphery

Paul Vicinanza[a], **Amir Goldberg**[a,1], **and Sameer B. Srivastava**[b,1]

[a]Stanford University, Graduate School of Business, 655 Knight Way, Stanford, CA 94305; ORCiDs: 0000-0003-3681-3380; 0000-0002-0858-3058; [b]University of California Berkeley, Haas School of Business, 2220 Piedmont Ave, Berkeley, CA 94720; ORCiD: 0000-0001-8793-0793

This manuscript was compiled on August 11, 2021

Where do visionary ideas come from? Although the products of vision as manifested in technical innovation are readily observed, the ideas that eventually change the world are often obscured. Here we develop a novel method that uses deep learning to identify visionary ideas from the language used by individuals and groups. Quantifying vision this way unearths prescient ideas, individuals, and documents that prevailing methods would fail to detect. Applying our model to corpora spanning the disparate worlds of politics, law, and business, we demonstrate that it reliably detects vision in each domain. Moreover, counter to many prevailing intuitions, vision emanates from each domain's periphery rather than its center. These findings suggest that vision may be as much as property of contexts as of individuals.

Vision | Contextual Word Embeddings | Deep Learning

Where do visionary ideas—such as those commonly attributed to Steve Jobs (1), Napoleon (2), Albert Einstein, or The Beatles—come from? This question has animated research across such diverse domains as scientific discovery (3–5), military strategy (6), and business (7, 8).

Because visionary ideas are difficult to observe, prior research has tended to focus either on the personal qualities of visionary leaders (9) or the contextual factors that enable vision to be realized in the form of concrete innovations—typically in the scientific or technical realm (3, 10, 11). This narrow empirical scope only allows for the identification of ideas that result in tangible inventions, predominantly in the form of patents or scientific publications. Yet most visionary ideas, from disruptive business strategies to paradigm-shifting legal interpretations, do not translate into singular innovations.

Moreover, vision does not merely inhere in novelty. Rather, visionary ideas are incommensurable, in the Kuhnian sense (12), with conventional logic. They fundamentally rethink the prevailing assumptions of the moment and later, as these assumptions evolve, gain widespread acceptance.

In the realm of politics, for instance, legislators regularly introduce and contest novel ideas that later become taken-for-granted assumptions. In the debates that raged about civil rights legislation in the U.S. during the 1960s, two of the staunchest opponents of these bills were Senators John Stennis and James Eastland. Although both voted against every major piece of civil rights legislation, they diverged in the nature of the opposition they put forward. Whereas Eastland—the least visionary senator according to our model—framed his opposition using overtly racist arguments (see SI and Fig. 1), Stennis, the senator deemed most visionary by our model, was among the first to base his objections on the principles of "color blindness," limited government, and individual freedom (13). This more subtle and indirect set of arguments would later become commonplace among opponents of civil rights legislation, laying the ideational bedrock for dominant contemporary conservative discourse on race relations in the US.

Existing methods, focused exclusively on the concrete products of innovation, would have been unable to detect traces of Stennis' vision. To address that, we develop a novel method, grounded in the tools of computational linguistics and machine learning, to identify visionary ideas independent of the form in which the idea is ultimately realized. Quantifying vision in this manner unearths prescient ideas that would otherwise be undetected by prevailing methods.

We measure vision using Bidirectional Encoder Representations from Transformers (BERT) (14), a deep neural network that encodes the semantic and contextual information of language. We exploit the ability of BERT to predict words given their context to first define *contextual novelty*: utterances that are poorly predicted by the model. To measure *vision*, we compare the contextual novelty at the time ideas are enunciated to contextual novelty at a later point in time. Ideas that are incommensurable with conventional logic at the time of their enunciation, but that become more commensurable in the future, are deemed visionary by our model.

Our novel method enables us to identify the locus of visionary ideas across a broad range of domains, addressing a fundamental question in the science of ideas and innovation: Where does vision come from? Received wisdom among stu-

Significance Statement

The most consequential changes in society—whether economic, artistic, legal, or political—stem from visionary ideas. Yet visionary ideas are difficult to directly observe, leading researchers to focus on how vision manifests in the form of domain-specific technical innovations. Thus, it has heretofore been impossible to consistently trace the origins of vision across domains. Here we develop a novel, deep-learning method to identify visionary ideas in language. We demonstrate that this common approach to quantifying vision can be applied across such disparate domains as law, politics, and business and that visionary ideas arise from the periphery rather than the center. Although vision is often attributed to individuals, its origins may have as much to do with place as with people.

[1]Amir Goldberg and Sameer B. Srivastava contributed equally to this work.

[2]To whom correspondence should be addressed. E-mail: pvicinan@stanford.com

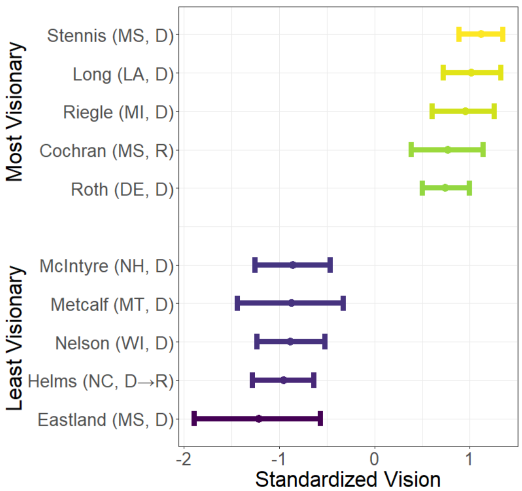

Fig. 1. Visionary Senators, minimum three congressional terms. Mean vision is computed, standardized, and bootstrapped 10K times at the politician-quarterly level.

dents of technological innovation is that transformative ideas emerge from the periphery because actors on the outskirts are less bound by institutional constraints and have greater freedom to explore new ideas (15–17). It is not obvious that this extends universally, however. In many non-technical domains it is commonly assumed that sage and established actors, such as Supreme Court judges (18), have the resources and experience to rethink existing logics. Even technological vision does not necessarily emerge from the periphery. Apple's groundbreaking iPhone and iPad, for example, were released when it was already among the world's most dominant technology companies. We apply our new method to three corpora that span the disparate domains of politics, law, and business, providing an unprecedentedly broad and multidisciplinary cartography of ideation. We demonstrate that visionary ideas are consistently recognized and rewarded by their relevant evaluators and that they emanate from the periphery rather than the center.

Defining Visionary Ideas

What makes an idea visionary? One key ingredient would appear to be novelty, as visionary ideas depart from the accepted conventions of the moment. Yet abundant prior research has shown that novelty does not by itself guarantee success (11, 19, 20). We propose that visionary ideas have two essential properties. First, they are novel in a particular way: they rethink the contextual assumptions that predominate a given field. By contextual assumptions, we mean those that: (a) are central to a domain's logics of action; and (b) guide a set of interdependent choices about how to configure activities for success in the field. In 1970, for example, Congress passed the Racketeer Influenced and Corrupt Organizations Act (RICO) to target organized criminal enterprises—in particular the Mafia. A small group of imaginative prosecutors soon seized upon RICO's ambiguous language to prosecute such wide-ranging civil crimes as mail fraud and stock manipulation (21). This approach was contextually novel in that it applied existing statutes intended for one set of actors to an entirely different class of actors and criminal activities.

Second, visionary ideas are prescient—that is, they foreshadow how the domain will evolve. While RICO's scope was initially limited, this approach to interpreting and applying RICO statutes well-beyond their original scope became commonplace among prosecutors and judges. It has since been used to prosecute organizations ranging from the Catholic Church to Major League Baseball to British Petroleum (BP) following the Deepwater Horizon oil spill (21). Notice that these two ingredients of vision—contextual novelty and prescience—can appear in an individual's discourse even when the person does not explicitly set out to predict the future, influence others, or even change the world. Moreover, visionary ideas can only be detected after the fact—that is, once the future state of the world is known.

Developing a Language-based Measure of Vision

Ideas are, of course, hatched by individuals and often expressed in discourse. The core idea of our approach is to measure the extent to which ideas expressed in routine discourse possess the quality of vision. To do so, we rely on the intuition that prescient ideas depart from prevailing ideas at the time of introduction but become commonplace in the future (5, 22). In particular, we use BERT, which learns the semantic and syntactic structure of language (in part) through a masked-word prediction task(14). BERT repeatedly predicts different masked (hidden) words in a sentence given the rest of the sentence, with the aim of minimizing the cross-entropy loss between the predicted and actual word. While most researchers apply BERT's model architecture to solve downstream tasks such as machine translation or text classification, we use the probabilistic features of the model to assess the extent to which a given set of ideas are visionary in their field.

Similar to how prior work trains separate word embedding models on a temporally split corpus to uncover semantic shifts in word meaning (23), by training separate BERT models over a split corpus, we reveal how the likelihood of specific words, phrases, and sentences evolves over time. Following standard practice, we begin with a pre-trained model and then fine-tune it to a given time interval (e.g., a year or presidential term) in each of our domains of interest.

To measure vision, we begin by considering perplexity, the exponentiated cross-entropy loss, which can be intuitively understood as the inverse-likelihood of the model generating a word or a document (normalized by the number of words). Higher perplexity scores correspond to unusual or unexpected utterances. We define *contextual novelty*, $CN(s)$, as the product of word-level perplexities in a sentence, s, normalized by the number of words in the sentence. We use mean sentence-level perplexity values to derive a measure of contextual novelty at the document level. *Vision* can then be operationalized as the percentage decrease in *contextual novelty* of a sentence between two time periods. Fig. 2 provides a schematic representation of our measurement approach, using a sentence from our legal data that was deemed highly visionary by our model. Depending on the context, we can then aggregate our measure of visionary discourse at different levels of analysis. In some settings, we can identify individuals who are apt to express visionary ideas. These individual-level measures can be aggregated to the level of social groups or organizations that might be more salient in other contexts. In still other

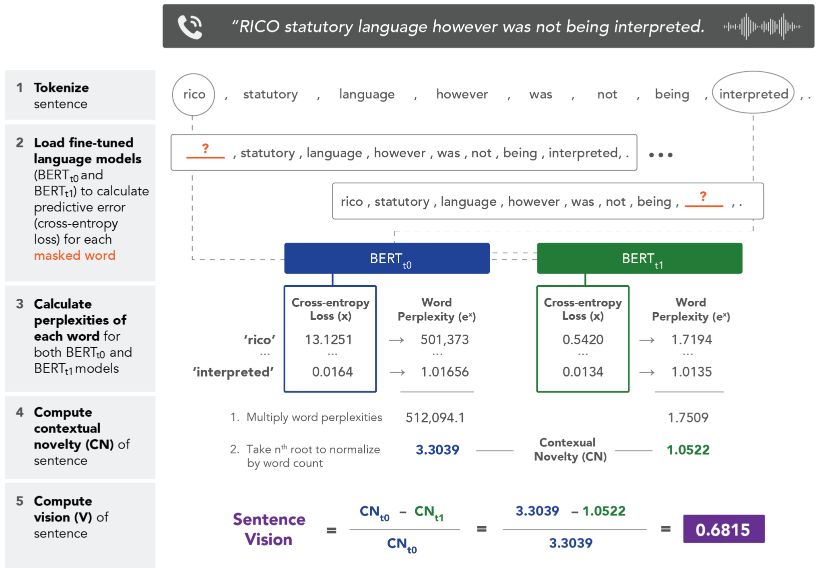

Fig. 2. Illustration of how vision is computed based on a sentence from the legal dataset that the model deems highly visionary. This sentence rates highly in vision due because the the RICO (Racketeer Influenced and Corrupt Organizations Act) token in the future period, when RICO' statutory language was heavily contested by the courts (21).

settings, the relevant unit of analysis might be a visionary document.

This approach to measuring vision offers several advantages over prior work. One feature is that any potential biases toward high perplexity sentences—such as rare tokens or errors in optical character recognition—are netted out in the numerator. Likewise, discussions unrelated to vision are netted out because the likelihood of a sentence must shift over time to result in a non-zero contribution to vision. Unlike topic models, our approach does not require tuning hyperparameters—though the researcher does have to make choices about how to partition the data. For more details please refer to the SI.

Results

Empirical Settings. We apply this method to identify visionary ideas in the domains of politics (4.9 million floor speeches given by members of the U.S. House of Representatives and the U.S. Senate), law (4.2 million rulings on U.S. State and Federal cases), and business (108,334 quarterly earnings calls in which the management teams of publicly traded firms lay out their vision and strategy for the company to financial analysts who cover their stock). Given that the corpora vary considerably in the time periods they cover and the nature of the discourse they include, we use slightly different approaches to fine-tuning BERT and defining the salient time horizon across the three (SI Appendix).

Model Validation. We assess the face validity of the words that our model identifies as most and least visionary in each of our three settings (SI, Table S1). In politics and law, we compared the early 1980s to the early 2010s. In both settings, the least visionary terms uttered in the early 1980s include ones related to the geopolitics of the Cold War between the U.S. and the U.S.S.R.—e.g., "MX (an intercontinental ballistic missile)" and "SALT (the Strategic Arms Limitation Talks)." The most visionary terms used in the early 1980s include ones that foreshadowed emerging health crises such as "HIV" and technological innovations such as "online."

In the corporate dataset, the most visionary term is "onboard," stemming from the token "onboarding." A business term originally referring to the assimilation of new employees, onboarding started gaining popularity in the early 2000s.[*] It was later reinterpreted to describe the integration of new users and customers onto a software platform and was only added to the Merriam-Webster dictionary in 2017. Firms whose leaders in 2011 talked about onboarding in this novel manner foresaw the rise of the Software as a Service (SaaS) business model.

Visionary Ideas are Rewarded. We next considered the relationship between vision and success in a given domain. Our analyses are based on the assumption that visionary ideas will generally be recognized and rewarded by the relevant audiences in a given field. Consistent with this expectation (Figure 3), vision is positively related to: a politician's likelihood of being reelected and her status attainment (Panels A and B; SI Table S2); a legal ruling's total citations and probability of being a landmark ruling (Panels C and D; SI Table S3); and a firm's annual stock returns (Panel E; SI Table S4). Indeed, it is only the highly visionary firms (top 5%) that achieve breakthrough levels of cumulative returns (Panel F).

The Origins of Visionary Ideas. We turn next to investigating the sources of visionary ideas. Positions in a given domain can be thought of as varying along a continuum from more central, which tend to be occupied by established actors that shape the rules and norms of a field, to more peripheral, which tend to be populated by upstart actors that face fewer institutional constraints (15).

Figure 4 shows that, across all three settings, truly visionary ideas—those at or above the 95th percentile of our continuous measure of vision—emanate from the periphery rather than the center. In politics, eigenvector network centrality (Panel A), K-core network centrality (Panel B), degree centrality, and closeness centrality are negatively related to the likelihood of a politician emerging as a true visionary (SI Table S6). In law,

[*] Based on the Google books n-gram viewer statistics at https://books.google.com/ngrams

lower (more peripheral) courts are more likely to produce truly visionary decisions than are upper (more central) courts (Panel C; SI Table S7). Estimating models with judge fixed effects that take advantage of the fact that judges are sometimes promoted across the judicial status hierarchy, we find that the likelihood of a judge authoring a visionary ruling declines by 0.8 percentage points after she is promoted from the U.S. District Court—the lowest rung of the federal judicial hierarchy—to the U.S. Appeals Court.[†] In business, as firm size—a proxy for centrality—based on total assets (Panel E) and number of employees (Panel F) increases, the likelihood of a firm being truly visionary declines (SI Table, S8).

Discussion

The ability to systematically quantify vision enables us to shed light on longstanding questions about the origins of transformative ideas. Popular intuitions often suggest that visionary ideas emerge from powerful incumbents. In law, for example, higher court judges are typically thought to produce visionary rulings which guide the subsequent judgments of lower-court judges (18). Similarly, in politics, established legislators who lead the most central committees are frequently identified as most visionary (25). In business, by contrast, theories of disruptive innovation implicitly assume that the visionary ideas underpinning revolutionary products and business models arise from new entrants to an industry rather than from entrenched incumbents (3, 26). Because we have heretofore lacked a systematic way of quantifying vision, such intuitions have been mostly informed by anecdotes and case studies. In contrast, our method reveals that, across a diverse array of domains, visionary ideas emanate from the periphery rather than the center.

Our results also suggest that vision may be as much as property of contexts as of individuals. Indeed, in the legal domain, the same individual becomes less visionary as she moves up the status hierarchy to more central courts. Those in search of breakthrough ideas should therefore look beyond the usual suspects who are ensconced in well-trodden places and instead focus attention on the unconventional ideas brewing in the outskirts of a domain.

Although our empirical investigation focused on three specific contexts, the method we introduce can be readily extended to detect vision in other domains. By focusing attention on and providing a novel means to quantify vision, we aim to broaden scholarly exploration from a narrow fixation on tangible artifacts such patents to the larger process by which visionary ideas that change the world emerge.

Materials and Methods

To extract vision from conversational text data, we build upon a recent innovation in natural language processing and deep learning: Bidirectional Encoder Representations from Transformers (BERT) (14). BERT is a generalized language model, meaning it learns the syntax and semantic meaning of language, which can then be

[†]Given the small number of justices in our data in the U.S. Supreme Court, our estimate is noisier and statistically indistinguishable from the probabilities for judges in U.S. District and Appeals courts.

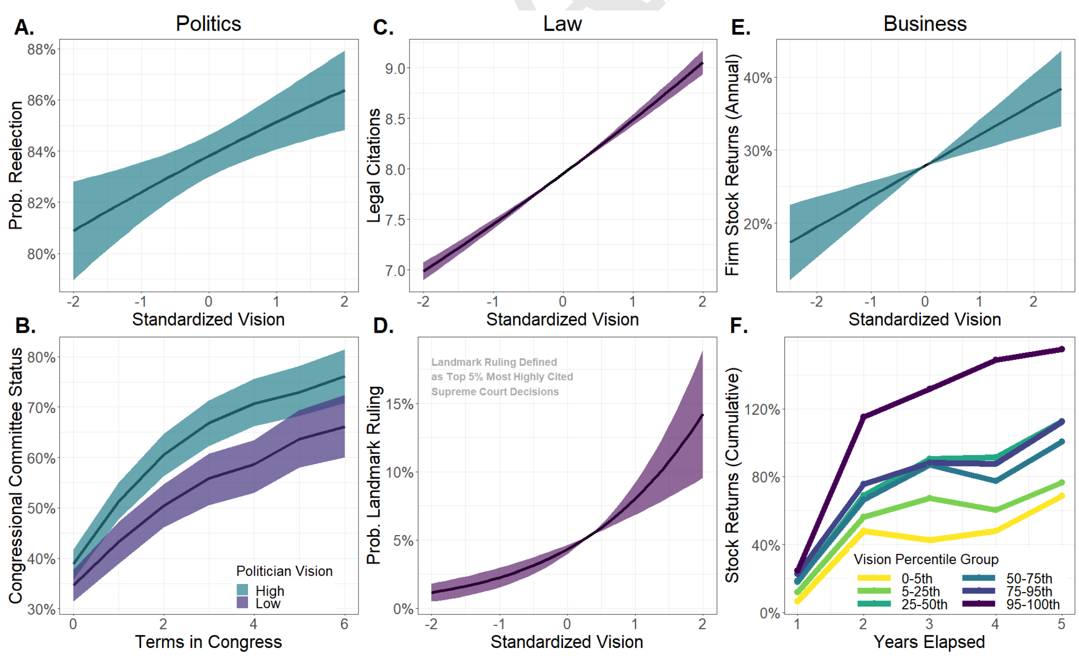

Fig. 3. Vision Predicts Success a) Vision predicts political re-election. Marginal effects plot from panel linear probability models of political reelection on standardized vision; politician-term unit of analysis, with political party x congressional term fixed effects ($\beta = 0.00860$, $p < 0.05$). **b) Vision predicts congressional committee status**. Mean committee status for the top tercile (high vision) and bottom tercile (low vision) Congressional term with 10K politician bootstrapped SEs. Congressional committee status defined by the committee transfer ratio (SI Appendix). Please see SI for panel regressions with fixed effects and other controls ($\beta = 0.0155$, $p < 0.001$). **c) Vision predicts highly cited court decisions**. Marginal effects plot of linear regression model of log total citations on standardized vision; judicial decision unit of analysis, with judge, court, and year fixed effects ($\beta = 0.0693$, $p < 0.001$). **d) Vision predicts landmark Supreme Court decisions**. Marginal effects plot of linear probability models of landmark decisions on standardized vision. Landmark decision is defined as the top 5% most highly cited U.S. Supreme Court decisions by year, with the sample restricted to U.S. Supreme Court decisions. Models include judge and year fixed effects ($\beta = 0.687$, $p < 0.001$). **e) Vision predicts firm stock returns**. Marginal effects plot of linear regression models of yearly stock returns from 2012-2015 on 2011 standardized prescience; NAICS 3-digit industry fixed effects ($\beta = 0.0422$, $p < 0.001$). **f) Vision predicts elite firm performance**. Total stock returns since 2012 by vision quartile and year (with top and bottom 5%). The y-axis shifts from annual stock returns in panel e to cumulative stock returns in panel f.

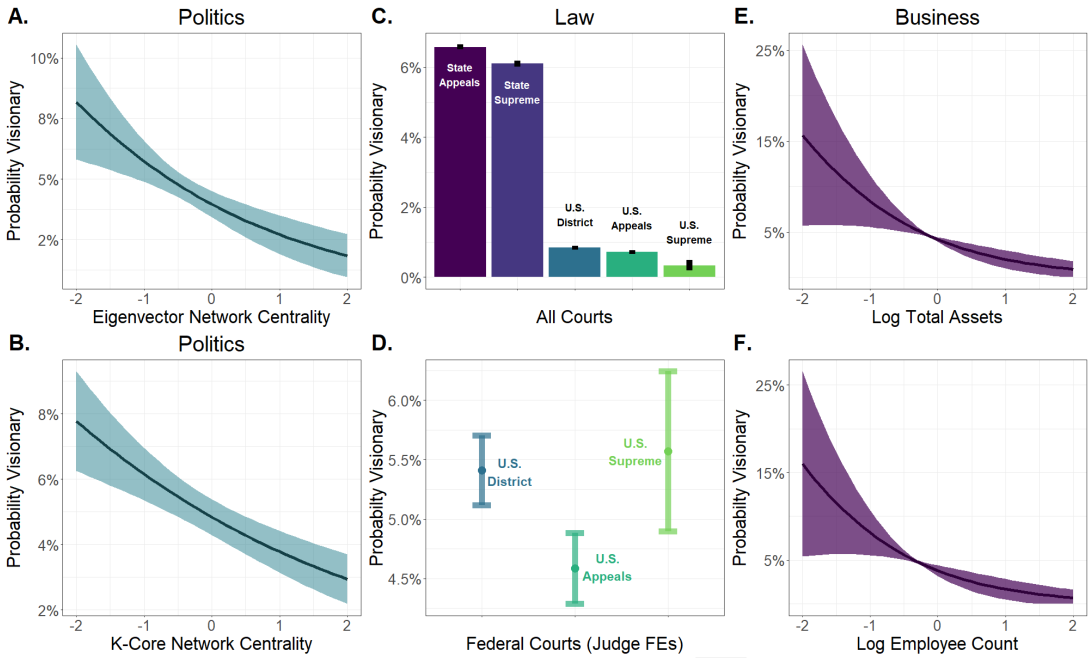

Fig. 4. Highly visionary ideas come from the periphery. Marginal effects plots regressing the probability of being highly visionary (top 5% in standardized vision) on alternatives measures of peripheral positions using logistic regression and 95% confidence intervals. All regressions include controls for the log number of sentences and are restricted to observations with at least 50 sentences given increased variance in vision with small sample size. **a & b) Highly visionary politicians come from peripheral network positions**. Politician network defined using bill consponsorship data (24). Network periphery measured by standardized eigenvector centrality ($\beta = -0.292$, $p < 0.001$) and standardized k-core centrality ($\beta = -0.277$, $p < 0.001$) with additional centrality measures in the SI. **c & d) Highly visionary court decisions come from the lower courts**. Panel C depicts the probability of a visionary decision using both state and federal courts and year fixed effects. Panel D adds judge fixed effects and restricts the sample to federal decisions (for which we have judge disambiguated decisions). **e & f) Highly visionary ideas come from small firms**. NAICS 2-digit industry fixed effects. Firm size measured by standardized total assets ($\beta = -.207$, $p < 0.01$) and the standardized number of employees ($\beta = -0.309$, $p < 0.05$).

applied to a litany of downstream tasks like machine translation and named entity recognition. Underlying BERT are layers of transformer blocks. The transformer architecture diverges from previous language modeling approaches by replacing recurrence and convolutions with attention mechanisms (27). Doing so allows the entire sentence to be propagated through the model simultaneously, significantly speeding up parallelization. BERT stacks dozens of transformer blocks, encompassing hundreds of millions of parameters. As a result, BERT learns syntax relationships, semantic meanings, co-references, and even encodes entire syntax trees (28, 29). Because BERT is computationally intensive—often requiring several weeks of time on dedicated cloud tensor processing units (TPUs) to train on a new corpus—researchers typically begin with the pre-trained model provided by Google (where BERT was developed) and fine-tune this model to their own corpora. Through the fine-tuning process, the general meanings learned by BERT can be contextualized to the researchers' specific domains of interest (30).

Traditional language models process sentences left-to-right, one word at a time, and estimate the conditional likelihood of a word: $(w_i|w_0, w_1, ...w_{i-1})$. BERT instead favors bidirectionality—that is, it attends to both the left and right contexts simultaneously (14). To circumvent the unidirectional constraint, BERT is trained (in part) using a masked language model (MLM) task: 15 percent of the words in a sentence are randomly masked, and the model is tasked to predict the masked tokens.

```
Sentence (s):  Earnings are up this quarter.
Masked s:  Earnings are [MASK] this quarter.
```

The MLM objective differs from "true" language models in that the likelihood of the model generating a sentence is undefined. As a proxy, we use the model's ability to solve the MLM for each word in the sequence, leaving all other words unmasked. Here, the likelihood of a word is conditional on both the left and right contexts: $(w_i|w_0, w_1, ...w_{i-1}, w_{i+1}, w_{i+2}, ...)$. While most researchers take the generalized language model features of BERT and add an additional layer on top to solve a downstream task, we instead directly use BERT's probabilistic modeling of language via MLM to quantify vision.

Specifically, we task the model to minimize the cross-entropy loss between the predicted and the actual word. Let $\mathbf{y}_i$ represent a location in a vector of length N, where N refers to the number of words in the corpus, for word i. This one-hot encoded vector takes a value of one at the index of the masked token and zero otherwise. $\hat{\mathbf{y}}_i$, also of length N, is the vector predicted token likelihood obtained through a softmax activation layer predicting the masked token by the BERT model. Model accuracy is evaluated using cross-entropy loss, $\mathcal{L}_i$ (Eq. 1). To obtain word-level perplexity, PP_i, which is the inverse-likelihood of the model generating the word, we exponentiate the cross-entropy loss.

$$\mathcal{L}_i = -\mathbf{y}_i \cdot \log(\hat{\mathbf{y}}_i) \quad [1]$$

$$PP_i = \exp(\mathcal{L}_i) \quad [2]$$

Words that are trivially predicted by the model—such as stop words and punctuation—have perplexities of approximating 1, meaning that the model predicts them with close to 100 percent accuracy. Conversely, words and phrases that are highly unusual or unexpected have higher perplexity scores. We take the product of these perplexities and normalize by the nth root to account for sentence length (Eq. 3). We refer to this term as contextual novelty (CN) instead of sentence-level perplexity for two reasons. First, given we use this measure to assess the extent to which ideas rethink the contextual assumptions in a domain, terming it contextual novelty aligns our empirical measure with our theoretical quantity of interest. Second, because BERT models bidirectionally, the perplexity of the sentence is technically undefined and terming this sentence-level perplexity would be inconsistent with prior work (31).

$$CN(s) = \left(\prod_{i=0}^{N} PP_i\right)^{\frac{1}{N}} \quad [3]$$

To measure *vision*, we rely on the intuition that visionaries' ideas depart from prevailing ideas at the time of introduction but become commonplace in the future (3, 32, 33). For example, Gerow et al. (22) identify highly influential scholarly publications by studying how academic discourse shifts after their publication. Thus, rather than fine-tuning BERT to our entire corpus, we split the corpus into time periods and fine-tune separate BERT models one each split of the corpus. This approach allows us to examine how the contextual novelty of a sentence changes over time.

Our method requires two BERT models trained on a corpus split into two periods, current (t_0) and future (t_1), and two BERT models which map documents to contextual novelties $CN_{t0}(s)$, $CN_{t1}(s)$. For a document from the current period, we define *vision*, V, as the percentage reduction in contextual novelty between the current and future models.

$$V(s) = \frac{CN_{t0}(s) - CN_{t1}(s)}{CN_{t0}(s)} \quad [4]$$

This approach to measuring vision offers several advantages over prior work such as topic models or TF-IDF vectorization. First, One feature is that any potential biases toward high perplexity sentences—such as rare tokens or errors in optical character recognition—are netted out in the numerator. Likewise, discussions unrelated to vision are netted out because the likelihood of a sentence must shift over time to result in a non-zero contribution to vision. Second, traditional pre-processing steps, such as removing stop words and punctuation, stemming, or converting to lowercase, are unnecessary in contextual embedding models. The only pre-processing of text prior to fine-tuning is sentence tokenization and appending `[CLS]` and `[SEP]` tokens to the start and end of each sentence respectively. BERT uses WordPeiece tokenization, which converts unrecognized tokens into sub-tokens (e.g., tokenizing `onboarding` into [`onboard`, `ing`] so no out-of-vocabulary words are dropped from the analysis.

Both *contextual novelty* and *vision* are defined at the sentence-level. To transition from the sentence level to the relevant unit of analysis in a given domain, we simply take the mean value of these variables over the unit of aggregation. For our corpus of U.S. State and Federal judicial decisions, for example, we aggregate at the unit of the judicial decision. We define truly visionary ideas (as manifested in individuals, organizations, or documents) as those in the top five percent of the distribution of vision. We find that our measure is noisier with short sentences, as there is less context for BERT to use when making predictions. To reduce noise, we restrict to sentences with at least 10 tokens and less than 100 tokens (to catch errors in the sentence parser) before aggregating to mean vision. Researchers replicating this methodology may consider using a higher minimum token count, such as 15 or 20 tokens, to further reduce noise.

Data, Fine-Tuning, and Measuring Vision. Training BERT from scratch is prohibitively expensive, taking weeks on a cloud TPU. Instead, Google has provided a pre-trained model—trained on the BookCorpus (800M tokens) and the English Wikipedia (2.5B tokens) available at https://github.com/google-research/bert—that researchers can fine-tune on their own corpora to learn context-specific idiosyncrasies. We fine-tune using BERT-Base uncased (12-layer, 768-hidden, 12-head, 110M parameter model) by repeating the MLM task and next-sentence prediction task on our corpora. We filter out sentences with less than 10 tokens to reduce noise and sentences longer than 100 tokens given that they likely represent errors in the sentence parser[‡]. For fine-tuning, we use a max sequence length of 128, a batch size of 64, and fine-tune for approximately 400,000 steps. Below we describe the three data sets in greater detail and explain the specific text pre-processing, fine-tuning, and approach to computing vision we followed in each setting.

[‡] An even higher minimum token count will greatly reduce noise in computing contextual novelty and vision.

Politics. To identify visionary politicians, we use transcripts from the United States House of Representatives and the United States Senate from the bound and daily editions of the United States Congressional Record from the 43rd to 114th Congresses (1873-2017). We use the data set provided by Gentzkow, Shapiro, and Taddy (2019), who remove procedural language and parse the text from each congressional session into speeches attributable to congresspeople (34). For data reliability reasons (e.g., temporal variation in optical character recognition (OCR) accuracy), we begin our analyses with the 87th Congress, whose members took office in 1961 resulting in 4.9 million unique speech events. We obtain biographical data on politicians from the Congressional Biographical Directory, GovTrack, and Congress.gov and collect committee membership (35, 36) and bill cosponsorship data (24).

We segment the corpus into four year increments, corresponding to presidential terms, resulting in 15 buckets—1961-1964, 1965-1968, ... 2007-2010—and fine-tune separate BERT models for each increment. To compute *vision*, we define the current period as the BERT model trained using the year the sentence was spoken, M_0. A more difficult choice is in selecting the future period. Choosing a proximate model quantifies short-term evolution in discourse, while a model trained on text further in time from the focal sentence quantifies longer-term vision. To balance this trade-off, we select two future period models—the immediately subsequent model M_1 and the model after that M_2—and take the arithmetic mean of vision between these two vision calculations: $[V(M_0, M_1) + V(M_0, M_2)]/2$. We define the current period model as the BERT model trained using the year the sentence was spoken. So, for example, for a sentence spoken in 1965, we use the BERT model trained on sentences from 1965 to 1968 as M_0 and the 1969-1972 and 1973-1976 models as the future models.

Law. Our data of U.S. State and Federal cases comes from the Caselaw Access Project, which digitized and processed over eight million state and federal judicial verdicts, stretching back to 1640s. To align these data with data on politicians and minimize OCR errors, we restrict our analysis to cases beginning in 1960. Our resulting sample includes 4.2 million cases. We obtain biographical data on federal judges, including their court tenure, gender, and prior judicial service from the Federal Judicial Center. Data on case citations come from the citation graph provided by the Caselaw Access Project, which extracts citations from the in-line text of court decisions. We remove in-line citations using `LexNLP`, a python package specifically designed to parse legal text, and sentence tokenize using this package as well. We compartmentalize the corpus into five-year intervals—1960-1964, 1965-1969...2005-2009—and fine-tune separate BERT models on each interval. As with the politics data set, we define the current period model as the BERT model trained using the focal year and compute mean vision using the subsequent two models to strike a balance between assessing shorter-term versus longer-term vision.

Business. To study visionary ideas in business, we collect a corpus of quarterly earnings calls (QECs) from seekingAlpha, a content service for financial markets. Our data set includes 108,334 QECs (414 million tokens) from publicly traded firms, predominately headquartered in the United States from 2006 and 2016. We restrict our analyses to the Q&A section of the call, removing the prepared remarks by the company and filtering out statements by analysts and the operator. We restrict to the Q&A portion because prepared statements have a very different style of discourse than the Q&A section which may complicate fine-tuning on a smaller corpus. We disambiguate company names in quarterly earnings calls and fuzzy match them to Compustat to obtain firm characteristics and performance outcomes. We identify 5,847 firms that have corresponding links to Compustat gvkeys. We then match gvkeys with Permno to link to the CRSP database, thereby allowing us to collect data on daily stock returns.

Unlike our datasets of political speeches and legal decisions, this corpus is comparatively small. We lack a significant number of transcripts for speeches between 2006 and 2011 and the text in the earlier speeches is heavily influenced by the 2007/2008 financial crisis. As a result, we restrict our analysis to QECs from 2011 onward. Given that these data span a relatively small number of years, we split the corpus on an annual basis and fine-tune a separate BERT model for each year in our data. We select 2011 as the focal year and select both 2015 and 2016, the last two years in our data, as the future comparison periods when computing vision. Because the unit of analysis is the firm, we aggregate all QECs for each firm in 2011 by computing the mean vision across calls in a given year to create a firm-level measure of vision.

Validating Vision. In Figure 3 of the main manuscript, we demonstrate how visionary actors reap rewards. For more details on these analyses and their associated regression tables please refer to the SI.

Politics. We begin by considering the extent to which vision appears to be rewarded for politicians in our data. We aggregate political vision at the congressional-term unit of analysis and select two measures of political success. The first, political *reelection*, is widely considered the primary motivator of incumbent politicians (37). The second variable, *committee status*, measures political success within the legislative chamber itself. We use the transfer ratio defined by Bullock and Spraque (1969), which is computed for each committee (38). We estimate linear regressions of reelection and committee status on average vision (Table S2).

Law. We turn next to considering whether vision is associated with success in the legal domain. We define success as the number of legal citations that accrue to the case. We restrict our analyses of citations to federal cases because doing so enables us to include judge-level controls, such as judge fixed effects, which are unavailable for state-level decisions. We also restrict to cases with at least 50 sentences to reduce variance in vision. This restriction drops less than one percent of decisions.

The first dependent variable of interest, *log citations*, is the natural log of the number citations a case receives plus one. We log transform the variable to reduce skew. A more stringent test of our method is whether it can identify landmark decisions—judicial rulings that significantly alter existing interpretations of existing law. Most landmark decisions come from the U.S. Supreme Court, given they have absolute authority to set precedent and determine law. We define a U.S. Supreme Court ruling as a *landmark decision* if it above the 95^{th} percentile in citations for the court. We estimate *log citations* using linear regression and *landmark decision* using logistic regression (Table S3).

Business. To validate our measure of vision in the business context, we began by testing whether our measure of vision can predict future stock returns. We measure stock returns using daily returns data: the percentage change in value accrued to the stockholder on a given day. We report the model looking at three-year returns from 2012 to 2014 in Figure 3, Panel E, although our findings are robust to all window (Table S5). We use 3-digit North American Industry Classification System (NAICS) fixed effects in all regressions to adjust for industry-level heterogeneity (Table S4 & S5).

Vision Arises from the Periphery of a Field. Next we turn to the core claim of the paper: visionary ideas are more likely to emerge from the periphery than from the center. There are many different ways to operationalize central versus peripheral positions—for example, based on an actor's position in the network of actors, relative size, and social status (e.g., based on demographic traits such as gender or elite credentialing). We use each of these different approaches in our three empirical settings. The choice of how to operationalize central versus peripheral positions is based on our understanding of each empirical context and the nature of the data that was available to us. The following section describes in detail the data and analyses used to create Figure 4 of the main paper.

We define highly visionary politicians, judicial decisions, and firms as the top five percent in vision during their relevant measurement period. For politicians this means top five percent in their current legislative term. We select the 95^{th} percentile and above as prior work predicting scientific impact uses this threshold (3, 11). When predicting *highly visionary*, it is essential to control for the number of sentences for each actor, because those with relatively few utterances are disproportionately likely to exhibit high levels of vision simply because our measure of vision is noisier with less data. We do so using *log sentence count*.

Politics. We define central and peripheral positions in the legislative body via politician's network position, as prior research demonstrates that well-connected politicians can influence their peers and governmental policy (24). To reconstruct the network of relationships between politicians we use data on bill cosponsorships (39–41) and construct separate bill consponsorship networks for each congressional term. We use a variety of network statistics to measure how peripheral a politician is in the cosponsorship network, including *degree centrality*, *eigenvector centrality*, *closeness centrality*, and *K-core centrality*. Because network centrality measures vary over time and between chambers, we standardize each network measure at the congressional term x chamber level. We regress *highly visionary politician* on our suite of network covariates using logistic regression (Table S6).

Law. In our legal setting, the U.S. Supreme Court is the most central because it has statutory authority over all other courts. Recognizing that jurisdictional differences render it difficult to make direct comparisons, we assume that state courts are more peripheral than federal courts because the fragmented body of state laws are consolidated by federal rulings but not vice versa. Thus, our judicial hierarchy goes from most peripheral to most central in the following order: State Appeals, State Supreme, U.S. District, U.S. Appeals, U.S. Supreme. We define *highly visionary decision* as the top five percent most visionary decisions in a given year across all courts. To test the idea that visionary judicial decisions come from the structural periphery of the court system, we regress *highly visionary decision* on separate indicator variables for each court using logistic regression (Table S7).

Business. Whereas periphery in the legal system is defined by hierarchy, we define periphery in business based on firm size. Large firms command significant market power and use their power to lobby congress for favorable regulations, set industry standards, and influence prices. By virtue of their size, large firms also occupy more central positions in the network between firms (42). We define firm size using *log total assets*, the total amount of economic value held by the firm and its number of employees, operationalized as *log employee count*. We define a *highly visionary firm* as one at or above the 95^{th} percentile in vision and estimate using logistic regression (Table S8).

ACKNOWLEDGMENTS. We thank Yixi Chen for invaluable research assistance in collecting and processing the data. We also thank Dan McFarland and Dashun Wang for helpful comments on prior drafts. The usual disclaimer applies. Data used in this study, including the fine-tuned BERT models, will be made publicly available post-publication.

1. W Isaacson, *Steve Jobs: The Exclusive Biography.* (Simon & Schuster), (2011).
2. A Roberts, *Napoleon: A Life.* (Penguin Books), Reprint edition edition, (2015).
3. L Wu, D Wang, JA Evans, Large teams Develop and Small Teams Disrupt Science and Technology. *Nature* **566**, 378–382 (2019).
4. J Singh, L Fleming, Lone Inventors as Sources of Breakthroughs: Myth or Reality? *Manag. Sci.* **56**, 41–56 (2010).
5. B Hofstra, et al., The Diversity–Innovation Paradox in Science. *Proc. Natl. Acad. Sci.* **117**, 9284–9291 (2020).
6. M Knox, W Murray, *The Dynamics of Military Revolution, 1300–2050.* (Cambridge University Press), (2001).
7. R Kanigel, *The One Best Way: Frederick Winslow Taylor and the Enigma of Efficiency.* (The MIT Press), (2005).
8. CM Christensen, *The Innovator's Dilemma: When New Technologies Cause Great Firms to Fail.* (Harvard Business Review Press), (1997).
9. MA Schilling, The cognitive foundations of visionary strategy. *Strateg. Sci.* **3**, 335–342 (2018).
10. S Wuchty, BF Jones, B Uzzi, The Increasing Dominance of Teams in Production of Knowledge. *Science* **316**, 1036–1039 (2007).
11. B Uzzi, S Mukherjee, M Stringer, B Jones, Atypical Combinations and Scientific Impact. *Science* **342**, 468–472 (2013).
12. TS Kuhn, *The Structure of Scientific Revolutions: 50th Anniversary Edition.* (University of Chicago Press), (2012).
13. JN Curtis, Remembering Racial Progress, Forgetting White Resistance: The Death of Mississippi Senator John C. Stennis and the Consolidation of the Colorblind Consensus. *Hist. Mem.* **29**, 134–160 (2017).
14. J Devlin, MW Chang, K Lee, K Toutanova, BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding in *Proceedings of NAACL-HLT*. pp. 4171–4186 (2019).

15. RK Merton, *The Sociology of Science: Theoretical and Empirical Investigations.* (University of Chicago Press), (1973).
16. SP Borgatti, MG Everett, Models of core/periphery structures. *Soc. networks* **21**, 375–395 (2000).
17. G Cattani, S Ferriani, A Core/Periphery Perspective on Individual Creative Performance: Social Networks and Cinematic Achievements in the Hollywood Film Industry. *Organ. Sci.* **19**, 824–844 (2008).
18. DA Schultz, CE Smith, *The Jurisprudential Vision of Justice Antonin Scalia.* (Rowman & Littlefield), (1996).
19. L Fleming, Recombinant Uncertainty in Technological Search. *Manag. Sci.* **47**, 117–132 (2001).
20. JG Foster, A Rzhetsky, JA Evans, Tradition and Innovation in Scientists' Research Strategies. *Am. Sociol. Rev.* **80**, 875–908 (2015).
21. L Coppola, N DeMarco, Civil RICO: How ambiguity allowed the racketeer influenced and corrupt organizations act to expand beyond its intended purpose. *New Engl. J. on Crim. Civ. Confin.* **38**, 241–256 (year?).
22. A Gerow, Y Hu, J Boyd-Graber, DM Blei, JA Evans, Measuring discursive influence across scholarship. *Proc. Natl. Acad. Sci.* **115**, 3308–3313 (2018).
23. N Garg, L Schiebinger, D Jurafsky, J Zou, Word embeddings quantify 100 years of gender and ethnic stereotypes. *Proc. Natl. Acad. Sci.* **115**, E3635–E3644 (2018).
24. JH Fowler, Connecting the Congress: A Study of Cosponsorship Networks. *Polit. Analysis* **14**, 456–487 (2006).
25. D Diller, S Stefani, *Richard G. Lugar: Indiana's Visionary Statesman.* (Indiana University Press), (2019).
26. C Christensen, RM McDonald, EJ Altman, J Palmer, Disruptive Innovation: Intellectual History and Future Paths. *Acad. Manag. Proc.* **2017** (2017).
27. A Vaswani, et al., Attention is All you Need in *31st Conference on Neural Information Processing Systems.* (2017).
28. J Hewitt, CD Manning, A Structural Probe for Finding Syntax in Word Representations in *Proceedings of NAACL-HLT.* pp. 4129–4138 (2019).
29. K Clark, U Khandelwal, O Levy, CD Manning, What Does BERT Look at? An Analysis of BERT's Attention in *Proceedings of the 2019 ACL Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP.* pp. 276–286 (2019).
30. Y Bengio, Deep Learning of Representations for Unsupervised and Transfer Learning in *Proceedings of ICML workshop on unsupervised and transfer learning.* pp. 17–37 (2012).
31. R Rosenfeld, Two decades of statistical language modeling: where do we go from here? *Proc. IEEE* **88**, 1270–1278 (2000).
32. B Kelly, D Papanikolaou, A Seru, M Taddy, Measuring Technological Innovation over the Long Run, (National Bureau of Economic Research), Working Paper 25266 (2020).
33. RJ Funk, J Owen-Smith, A Dynamic Network Measure of Technological Change. *Manag. Sci.* **63**, 791–817 (2016).
34. M Gentzkow, JM Shapiro, M Taddy, Measuring Group Differences in High-Dimensional Choices: Method and Application to Congressional Speech. *Econometrica* **87**, 1307–1340 (2019).
35. C Stewart III, Committee Hierarchies in the Modernizing House, 1875-1947. *Am. J. Polit. Sci.* **36**, 835–856 (1992).
36. C Stewart III, J Woon, Congressional Committee Assignments, 103rd to 114th Congresses, 1993-2017 (2017).
37. JA Schlesinger, *Ambition and Politics: Political Careers in the United States.* (Rand McNally and Co, Chicago, IL), (1966).
38. C Bullock, J Sprague, A Research Note on the Committee Reassignments of Southern Democratic Congressmen. *The J. Polit.* **31**, 493–512 (1969).
39. JH Fowler, Legislative cosponsorship networks in the US House and Senate. *Soc. Networks* **28**, 454–465 (2006).
40. L Ten Brinke, CC Liu, D Keltner, SB Srivastava, Virtues, vices, and political influence in the us senate. *Psychol. Sci.* **27**, 85–93 (2016).
41. CC Liu, SB Srivastava, Efficacy or rigidity? power, influence, and social learning in the us senate, 1973–2005. *Acad. Manag. Discov.* **5**, 251–265 (2019).
42. MP Allen, The Structure of Interorganizational Elite Cooptation: Interlocking Corporate Directorates. *Am. Sociol. Rev.* **39**, 393–406 (1974).