WHO SEES THE FUTURE? A DEEP LEARNING LANGUAGE MODEL DEMONSTRATES THE VISION ADVANTAGE OF BEING SMALL

A PREPRINT

Paul Vicinanza
Stanford University
Graduate School of Business
pvicinan@stanford.edu

Amir Goldberg
Stanford University
Graduate School of Business
amirgo@stanford.edu

Sameer B. Srivastava
University of California, Berkeley
Haas School of Business
sameersriv@berkeley.edu

May 26, 2020

ABSTRACT

Which groups are most likely to become visionaries that define the future of their field? Because vision is difficult to measure, prior work has reached conflicting conclusions: one perspective emphasizes the benefits of being large, established, and central, while another stresses the value of being small, upstart, and peripheral. We propose that this tension can be resolved by disentangling vision—the capacity to generate contextually novel ideas that foretell the future of a field—from the traces of vision that result in tangible innovation. Using Bidirectional Encoder Representations from Transformers (BERT), we develop a novel method to identify the visionaries in a field from conversational text data. Applying this method to a corpus of over 100,000 quarterly earnings calls conducted by 6,000 firms from 2011 to 2016, we develop a measure—prescience—that identifies novel ideas which later become commonplace. Prescience is predictive of firms' stock market returns: A one standard deviation increase in prescience is associated with a 4% increase in annual returns, and firms exhibiting especially high levels of prescience (above the 95$^{\text{th}}$ percentile) reap especially high returns. Moreover, contrary to theories of incumbent advantage, we find that small firms are more likely to possess prescience than large firms. The method we develop can be readily extended to other domains to identify visionary individuals and groups based on the language they use rather than the artifacts they produce.

Keywords Innovation · Contextual Word Embeddings · Firm Performance · Vision

1 Introduction

What sets visionaries such as Steve Jobs [1], Napoleon [2], and The Beatles apart from their contemporaries? This question has animated research across domains as diverse as scientific discovery [3, 4], art [5], fashion [6], and business strategy [7].

Whereas an extensive literature has focused on the factors that lead certain individuals to be visionaries in their field [1], here we focus on the types of groups that are more or less likely to produce visionary ideas. Prior work has reached conflicting conclusions on the relationship between group size and visionary advantage. One perspective suggests that visionary ideas are most likely to emanate from large, established, and central groups because they have: (a) the resources needed to accurately forecast how the future will unfold [8]; (b) the power and influence to shape the choices of other actors in ways that conform to their vision of the future [9]; and (c) the status that will compel other actors to want to emulate their choices [10, 11]. An alternative view proposes that small, upstart, and peripheral groups are more likely to have vision because they have: (a) greater freedom to deviate from normative expectations [12]; (b) less investment in maintaining the status quo [13, 7]; and (c) greater risk tolerance for experimenting with novel ideas [14, 15].

Empirical evidence in support of these competing perspectives has been decidedly mixed [16]. We trace this lack of consensus to the fact that prior research has mostly focused on innovative artifacts such as patents that arise from technical innovation and that are readily observed. Vision does not, however, necessarily result in technological innovation [7]. Rather, it inheres in the difficult-to-observe ideas that originate within groups and that fundamentally restructure how a field will operate in the future. The Oakland Athletics' surprising success in 2002, for example, was a result of their integration of data analytics into the process by which they assembled their team and their defiance of conventional wisdom in scouting, developing, and trading players [17]. A small and chronically underfunded baseball team, the Athletics outperformed established and powerful competitors, eventually breaking a 100-year-old record with 20 consecutive victories. Indeed, many examples of visionary breakthroughs—from Amazon's redefinition of retail [18] to Mohandas Gandhi's nonviolent resistance of British occupation—are not rooted in the production of tangible artifacts amenable to scientific inquiry.

What, then, makes some ideas more visionary than others? One key ingredient would appear to be novelty. Yet abundant prior research has shown that novelty does not by itself guarantee breakthrough success [19, 20, 21]. We instead 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 group's logic of action; and (b) guide a set of interdependent choices about how to configure a group's activities for success in the field [22]. In Steve Jobs' words, visionary ideas "connect the dots" in novel ways.[1] The Oakland Athletics, for example, developed new ways of assessing player quality, which led them to fundamentally redesign the process by which they scouted, developed, and traded for players. These choices fundamentally altered the Athletics' strategic position in the field of baseball. Second, visionary ideas are prescient—that is, they foresee how the field will evolve. The Oakland Athletics' sabermetric-based approach was soon emulated by other teams, fundamentally reshaping baseball's competitive landscape [17]. To put it differently, a group is visionary when its ideas question the contextual assumptions of a field in ways that peers fail to see in the moment but that they come to adopt in the future. Groups can be visionary because they are skilled at forecasting the future, because they influence others to follow their lead, or some combination of both.

Extending Bidirectional Encoder Representations from Transformers (BERT) [23], we develop a novel measure that identifies from conversational text when groups produce ideas that are both contextually novel and prescient. Groups that rank high on this measure, we contend, are the visionaries in their field. BERT, a multi-layer neural net model, is particularly effective at representing semantic and contextual information. We exploit this property to identify the forms of novelty that are most likely to lead to a vision advantage: utterances that the model identifies as contextually surprising, or that "connect the dots" in novel ways. We refer to this property as *contextual novelty*. To measure the prescience of such utterances, we compare the contextual novelty at the time of enunciation to contextual novelty at a later point in time. We refer to this property as *prescience*. Prescient groups are those that utter statements that are novel in the moment but become commonplace in the future.

Our method departs from existing approaches in two fundamental ways. First, prior work mostly focuses on the products of innovation—e.g., patents [3], scientific articles, and songs [24]—and can therefore only detect the traces of vision that manifest in these artifacts. Our approach instead identifies visionary narratives from conversational text. Second, because it directly models context, BERT enables us to approximate the logic informing speakers' utterances. Unlike bag-of-words textual analysis methods that are based on the co-occurrences of words [25], BERT encodes information about entity types and their relationships in a manner that represents conceptual knowledge about the world [26]. Deviations from this model correspond to novel (but not necessarily correct) understandings of reality. Such deviations that are also prescient can be thought of as visionary.

To demonstrate this method's utility, we apply it to a corpus of over 100,000 quarterly earnings calls conducted by over 6,000 publicly traded firms from 2011 to 2016. These firms constitute a significant portion of the world economy, with annual revenues exceeding 75% of U.S. GDP. We use our method to identify firms that possess vision—a quality that is frequently invoked in organizational and economic research but has not been systematically measured. We find that firms high in prescience—namely those that possess vision—are more successful several years in the future: they exhibit higher stock returns, greater earnings, and more robust growth. Moreover, our method enables us to identify the types of firms that possess a vision advantage. Contrary to theories of incumbent advantage, we find that small firms are more likely to be visionary than large firms.

Measuring Vision through Language Use

Our measures of contextual novelty and prescience are based on the tools of natural language processing (NLP) and deep learning—in particular, the BERT model [23]. BERT learns the semantic and syntactic structure of language (in

[1] https://news.stanford.edu/2005/06/14/jobs-061505

part) through a masked-word prediction task: it repeatedly predicts different masked 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. 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.

While most researchers apply the model architecture to solve downstream tasks such as machine translation or text classification, we use the probabilistic features of the model to quantify the extent to which actors are visionaries in their field. 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) [27]. BERT computes perplexity for individual words via the masked-word prediction task. Words that are readily anticipated—such as stop words and idioms—have perplexities close to 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 define *contextual novelty*, $PP(s)$, as the product of word-level perplexities in a sentence s normalized by the number of words in the sentence. We use mean sentence perplexity values to aggregate perplexity to the document level.

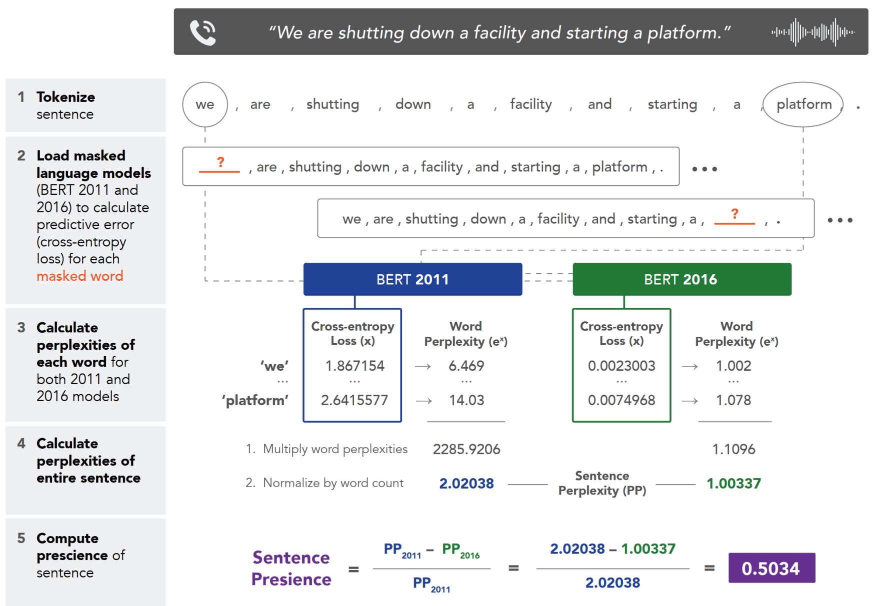

Figure 1: Example prescience calculation for a highly prescient sentence from our data

To measure *prescience*, we rely on the intuition that visionaries' ideas depart from prevailing ideas at the time of introduction but become commonplace in the future [28, 3, 29]. For example, Gerow et al. [30] 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 fine-tune BERT to each year in our data. This approach allows us to examine the extent to which the perplexity of a sentence uttered in one year changes in a different year.

Just as word embedding models trained on a temporally split corpus uncover semantic shifts in word meaning [31], our approach reveals how the contextual novelty (i.e., perplexity) of a sentence or document changes over time. We define change in contextual novelty from one period to another as prescience. In the simplest form of the measure, assume a corpus of documents, $i \in I$, that is split into current and future periods, I_c and I_f, and two BERT models that map documents to perplexities $PP_c(i)$, $PP_f(i)$. Language that is more typical in the future period than in the present period will have a lower perplexity in the model trained on future documents, $PP_c(i) > PP_f(i)$. For a document from the current period, we define *prescience* as the percent reduction in perplexity between the present and future models: $(PP_c(i) - PP_f(i))\ /\ PP_c(i)$. One feature of this approach 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. Importantly, 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 prescience. Unlike topic models [32], our approach does not require tuning hyperparameters—though the researcher does have to make choices about how to partition the data. Moreover, because our measure of prescience is calculated at the sentence level, it is not affected by other co-occurring sentences as would be the case with bag-of-words approaches. Fig. 1 provides a schematic representation of our measurement approach, using a highly prescient sentence from our data.

Results

Setting

We evaluated our model of prescience in the field of for-profit firms. To ascertain vision, we collected a corpus of over 100,000 quarterly earnings calls (QECs) of publicly traded firms. The top management teams of most publicly traded firms hold QECs with financial analysts who make recommendations on whether investors should buy, sell, or hold the stocks of firms they cover. QECs are consequential because they provide corporate leaders with an opportunity to publicly discuss and put into context their past performance, lay out their vision and strategy for the future, and address analysts' questions and concerns (SI Appendix) [33].

Although we had access to data going as far back at 2006, we opted to begin our analyses in 2011 to circumvent the potentially confounding effects of the 2008 financial crisis. We fine-tuned BERT to each year between 2011 and 2016. In the analyses presented below, we focus on prescience as defined by changes in perplexity between 2011 and 2016. We obtained comparable results when we choose different windows of time (see Methods section).

Model Validation

Scholars of innovation have long noted that novelty comes from the periphery [12]. Large firms neglect novel technological opportunities in favor of improving existing product lines [13], just as large scientific teams build upon recent high impact work and reap immediate rewards [3]. Building on these insights, we sought to validate our measure of contextual novelty by examining whether small firms in our data exhibit greater contextual novelty than large firms. Measuring firm size as the natural log of total assets, we found strong support for this assertion ($p < 0.00001$). A 1 SD decrease in firm size corresponds to a 0.27 SD increase in contextual novelty. Because the theoretical arguments that suggest smaller firms will be more contextually novel than larger firms can also be extended to the case of lower status firms relative to higher status firms, we conducted a supplemental analysis examining how status relates to contextual novelty. Using securities analyst coverage as a proxy for status [34], we found that a 1 SD decrease in firm status corresponds to a 0.31 SD increase in contextual novelty ($p < 0.00001$).

To validate our model, we begin by inspecting words that are used least and most presciently in our data. Among the least presciently used words are terms such as `DVD` (lowest prescience token), `CD`, `disk`, and `PCs`, likely reflecting the shift from physical media to cloud storage, as well as words relating to major events in 2011 such as `stimulus`, `tsunami`, `Greece`, `deficit`, and `Iraq`.

Prescient word usage, in contrast, relates to visionary business strategies (SI Appendix, Table S1). The most prescient word in our dataset—twice as presciently used as any other 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[2] and was added to the Merriam-Webster dictionary in 2017. As the software-as-a-service (SaaS) business model was becoming prevalent in the 2010s, this term was reinterpreted to describe the integration of new users and customers. Firms talking about onboarding in this context in 2011 foresaw this evolution. Similarly, among the 5 most prescient sentences is a reference to `compassionate use pre-approvals`, or the use of a novel unapproved drug to treat seriously ill patients when no other treatments are available (SI Appendix, Table S2). Although legally fraught and historically very rare, rates of FDA compassionate use pre-approvals more than doubled between 2008 and 2016.[3]

Along the same lines, we can identify firms that are most and least prescient in 2011. The most prescient list includes mostly mid-sized firms that operate in knowledge-intensive sectors such as medical devices (e.g., SurModics), information technology (e.g., Openwave Systems), clean energy (e.g., Xcel Energy), fiber optics (e.g., EMCORE) and pharmaceutical drug development (e.g., Neurocrine Biosciences). The least prescient firm, Oil-Dri Corporation, has been in existence for over 75 years and makes products from sorbent minerals (e.g., cat litter). Other low prescience firms include beverage packaging (Crown Holdings) and energy firms (CONSOL Energy) that were founded in the 19th century, as well as Yingli Green Energy that announced in 2018 that it would be delisted from the New York Stock Exchange. Yet there are also some surprises on the least prescient firm list such as Glu Mobile (a mobile gaming firm) and Netflix. Though quirky, the Netflix example is quite revealing about what the prescience measure picks up on. In 2011, Netflix announced it would split its mail order DVD service into a separate entity. Seen as backward-looking by investors, the move resulted in an almost 80% drop in Netflix's stock price, leading it to backtrack. Though two years

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

[3]FDA Expanded Access submission statistics reported at https://www.fda.gov

Table 1:

Most Prescient Firms	Least Prescient Firms
SurModics, Inc.	Oil-Dri Corporation of America
Openwave Systems Inc.,	Rovi Corporation
Methode Electronics, Inc.	Crown Holdings, Inc.
Xcel Energy Inc.	Glu Mobile, Inc.
Neurocrine Biosciences, Inc.	CONSOL Energy Inc.
Polaris Industries, Inc.	Netflix, Inc.
CenterPoint Energy, Inc.	Yingli Green Energy
Incyte Corporation	Asure Software, Inc.
Crosstex Energy	Energen Corporation
EMCORE Corporation	Hasbro, Inc.

The most and least prescient firms: Minimum 50 sentences.

later Netflix would pioneer an online content production model, in 2011 it struggled to break from its past and develop a coherent strategy. Overall, Table 1 provides further validation that our measure of prescience uncovers the linguistic markers of vision.

The Downstream Correlates of Prescience

Assuming that visionary firms can capitalize on their foresight, we hypothesize that visionaries will also enjoy greater marketplace success in the future relative to their peers. Figure 2, Panel A reports marginal effects from an ordinary least squares regression estimating future returns to stock ownership. The dependent variable is total returns to stock ownership from the start of 2012 onward. Total returns, as opposed to raw stock price, accounts for dividends, stock splits, and other stock manipulations that influence shareholder returns [35]. The key independent variable is 2011 prescience, and the unit of analysis is the firm ($n = 1,027$). We include 3-digit NAICS industry fixed effects to account for industry-level variation in returns. We find that a 1 SD increase in prescience corresponds to a 4-percentage point increase in annual returns to stock ownership ($p < 0.001$) (SI Appendix, Table S3).

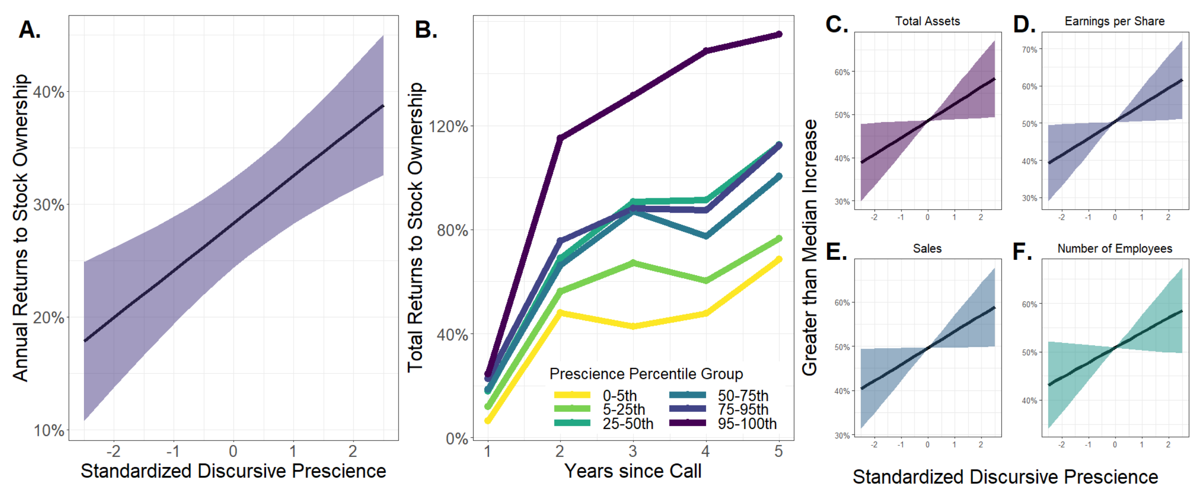

Figure 2: **Prescience is positively related to firm performance:** All 1,031 firms with QECs in 2011 and at least 50 sentences. **a)** Marginal effects plot of yearly stock returns from 2012-2015 on 2011 standardized prescience. NAICS 3-digit industry fixed effects and robust clustered standard errors. 95% confidence intervals computing using delta-method approximation. **b)** Total stock returns since 2012 by prescience quartile and year (with top and bottom 5%). The y-axis shifts from annual stock returns in **a** to cumulative stock returns. **c-f) Prescience is positively related to other indicators of firm performance:** Marginal effects plots of the percentage change in each firm metric between 2012 and 2015 on standardized prescience. Because variables are non-linear, skewed, and take negative values, we dichotomize our DVs as above or below the median value. textbfc) $\beta = 0.177$, $p < 0.05$. **d)** $\beta = 0.204$, $p < 0.05$. **e)** $\beta = 0.188$, $p < 0.05$. **f)** $\beta = 0.140$, $p < 0.1$

This average effect obscures strong non-linearities in the relationship between prescience and firm returns. Figure 2, Panel B depicts the relationship between market performance and prescience by percentile group. We find little difference in stock returns between the 25^{th} and 90^{th} percentile in prescience. However, firms in the top 5 percent have exceptionally high market returns: 50 percentage point higher stock returns than average just 3 years after 2011 (SI Appendix, Table S4).

In addition to total stock market returns, we evaluated our measure against other forms of firm performance. Figure 2 Panels C, D, E, and F report the marginals effects of growth in assets (logged), earnings per share, sales, and the number of employees on prescience. A 1 SD increase in prescience corresponds to approximately a 5% increase in the likelihood a firm experiences above average growth in all of these categories. These results suggest that possessing a vision advantage confers tangible rewards to a firm (SI Appendix, Table S5).

The Vision Advantage

What kinds of firms possess a vision advantage? To begin, we find significant industry variation in prescience (SI Appendix, Fig. S1). Firms in mining, utilities, and construction are least prescient, while those in the professional services sector (e.g., finance and insurance, information technology, management consulting) are most prescient. In general, prescient firms appear to be concentrated in knowledge-intensive industries where the traditional artifacts of innovation (e.g., patents) are not consistently relevant.

Second, we find no relationship between prior firm performance (as measured by return on assets, Tobin's Q, or prior stock returns) or technological innovation (as measured by patenting success) and prescience. In other words, highly prescient firms do not: spend more on R&D, disproportionately come from high-tech industries, patent at higher rates, or produce higher impact or disruptive patents (SI Appendix, Table S6) [36]. Given the concentration of prescience in non-patenting industries, it suggests our measure of vision, applied to QECs, quantifies the non-technical aspects of business innovation.

Finally, we evaluate whether smaller firms are more likely to possess a vision advantage over larger firms. Prior work reports that the effects of novelty accrue to those at the top of the distribution [3, 20]. Building on this insight as well as our finding that the advantages of prescience are concentrated at the top of the distribution (Fig. 2B), we estimate logit models of "elite" prescience (i.e., whether or not a firm is in the top 5% of the distribution of prescience) on firm size and include industry fixed effects (using 2-digit NAICS codes). Our results are robust to alternative cut-points in the definition of elite prescience.

We operationalize firm size in two ways: total assets and number of employees. These two measures' relationships with prescience are illustrated in Panels A and B of Fig. 3, respectively. Results for both models are highly significant ($p = 0.001$) and substantive. A one standard deviation decrease in firm size relative to the mean doubles the likelihood of exhibiting elite prescience. A firm 1 SD below the mean in assets (employees) is five (four) times more likely to exhibit elite prescience than a firm 1 SD above the mean (SI Appendix, Table S7).

Discussion

Popular accounts of innovation tend to celebrate change that comes from upstart outsiders. Yet visionary thinking, such as Apple's introduction of tablet computing [1] or the late twentieth century Revolution in Military Affairs promoted by the US [37], often originates in large, established, and previously trailblazing players. It is tempting to assume that groups that have been successful in the past by virtue of their superior resources, power, or status will be the fountains of visionary ideas. Yet the relationship between group size and the production of visionary ideas has heretofore been obscured because prior research has been limited to studying the products of innovation rather than visionary ideas themselves [3]. We take advantage of the possibilities afforded by advances in natural language understanding and, using BERT, develop a measure of vision in conversational text. Our method provides a way to analytically distinguish visionary ideas from the traces of those ideas that result in tangible outputs such as patents. Consistent with popular intuitions, and contrary to theories of incumbent advantage, we demonstrate that smaller and less established firms are more likely to be visionaries than their larger and more entrenched peers.

Our empirical investigation focused on for-profit firms, but our method can be readily extended to detect vision in other domains—from identifying political vision in speeches and debates to tracing visionary arguments in legal proceedings to detecting visionary knowledge in scholarly publications. By focusing attention on and providing a new way to measure vision, we aim to broaden scholarly exploration from a narrow fixation on the tangible artifacts of innovation to the broader process by which individuals and groups conceive of ideas that are contextually novel and prescient. It is these ideas that ultimately change the world.

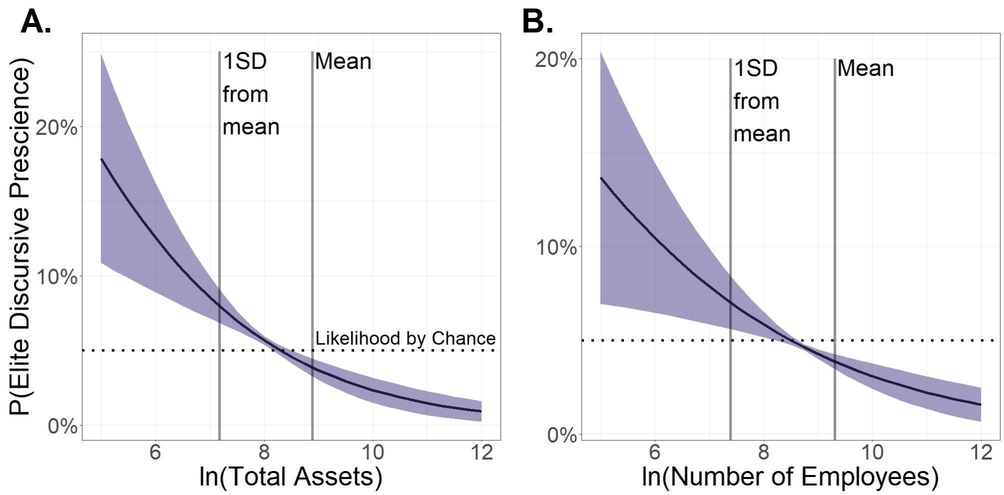

Figure 3: **Firm size is negatively associated with elite prescience** Marginal effects plot from logistic regression predicting elite prescience (top 5%). NAICS 2-digit fixed effects and robust clustered standard errors. with 95% confidence intervals computing using delta approximation. Controls for number of calls and number of sentences. All firms with QECs in 2011, minimum 100 sentences. **a)** Firm size as measured by total assets ($\beta = -.489$, $p = 0.000$). **b)** Firm size as measured by number of employees ($\beta = -.350$, $p = 0.001$)

2 Materials and Methods

Bidirectional Encoder Representations from Transformers

(BERT) has achieved state-of-the-art performance on numerous natural language processing (NLP) tasks, including question answering, entailment, and semantic equivalence [23]. While the model progresses several innovations simultaneously, one key innovation is the bidirectional nature of the language model. To train bidirectionally, the authors use a masked language model objective (MLM), where the model is tasked to predict a series of randomly masked words in a sentence.

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

In evaluation, BERT estimates token likelihood through a softmax activation layer. We evaluate likelihood through cross-entropy loss ($\mathcal{L}$) and exponentiate to obtain word-level perplexities (PP). For a masked word i in sentence s:

$$\mathcal{L}_i = -\vec{y}_i \cdot \log(\vec{\hat{y}}_i)$$
$$PP_i = \exp(\mathcal{L}_i)$$

The MLM objective differs from "true" language models in that the likelihood of the model generating 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. We take the product of these perplexities and normalize by the nth root to account for sentence length. We term this averaged word-level perplexity but acknowledge it's qualitative differences with sentence perplexity in a traditional language model (See SI Appendix for more details) .

$$PP(s) = \left(\prod_{i=0}^{s} PP_i\right)^{\frac{1}{N}}$$

Measuring Vision

We operationalize vision using two measures, *contextual novelty* and *prescience*. Both measures are defined at the sentence-level and aggregated by averaging sentences across the actor. Contextual novelty is defined as the raw perplexity of the sentence, $PP(S)$.

To measure prescience, our method requires two BERT models trained on a corpus split into two periods, current and future: $I_c \in I$ and $I_f \in I$ and two BERT models which map documents to perplexities $PP_c(i)$, $PP_f(i)$. For a document from the current period, we define *prescience* as the percentage reduction in perplexity between the current and future models:

$$\text{Prescience} = \frac{PP_c(i) - PP_f(i)}{PP_c(i)}$$

Robustness Check

To provide robustness, we compute prescience using alternative years of QECs. Using 2015 instead of 2016 as the future year we strongly replicate our findings. A 1SD increase in prescience corresponds to a 4 percentage point increase in annual stock returns ($p < 0.001$). Using 2012 as the current year, instead of 2011, and keeping 2016 as the future year also reproduces the relationship with stock returns ($p < 0.01$, 3.5 percentage point increase) but only for the first year after the call. Note that the QECs come from 2012 and the sample is entirely different from our main analyses, yet we obtain identical marginal effects. The relationship between firm size and elite prescience in both models is significant without controls ($p < 0.001$) and remains in the proper direction but loses significance when controls are included.

One concern may be that we "test" the model on the same data we use to train it. To address this concern, we train the current year of the model using 2010 QECs (leaving the future year unchanged) and evaluate our model of prescience on the 2011 QECs. In this modeling approach, we evaluate prescience on an entirely new set of QECs. We find that prescience predicts future returns, but only from the time period between 2012 to 2013 ($p < 0.01$). A 1 SD increase in prescience corresponds to a 3.4 percentage point increase in stock returns. The relationship between firm size and elite prescience is in the proper direction but no longer significant. These robustness analyses suggest that any gains to prescience are short-lived as the market quickly adjusts. To measure long-term prescience we likely need data which stretches longer than 6 years.

Data Access

Data used in this study, including the fine-tuned BERT models, will be published post-publication.

Acknowledgments

We thank seminar participants at the Berkeley TextXD conference, the Berkeley Culture Conference, and the Berkeley Open Innovation Seminar. We would also like to thank Chris Potts, Michael Hahn, Dan Jurfasky, and Daniel McFarland for their feedback and advice on this project.

References

[1] Melissa A. Schilling. *Quirky: The Remarkable Story of the Traits, Foibles, and Genius of Breakthrough Innovators Who Changed the World*. PublicAffairs, February 2018.

[2] Andrew Roberts. *Napoleon: A Life*. Penguin Books, reprint edition edition, October 2015.

[3] Lingfei Wu, Dashun Wang, and James A. Evans. Large teams Develop and Small Teams Disrupt Science and Technology. *Nature*, 566(7744):378–382, February 2019.

[4] Jasjit Singh and Lee Fleming. Lone Inventors as Sources of Breakthroughs: Myth or Reality? *Management Science*, 56(1):41–56, January 2010.

[5] Brian Uzzi and Jarrett Spiro. Collaboration and Creativity: The Small World Problem. *American Journal of Sociology*, 111(2):447–504, September 2005.

[6] Diana Crane. Globalization, organizational size, and innovation in the French luxury fashion industry: Production of culture theory revisited. *Poetics*, 24(6):393–414, July 1997.

[7] Clayton M. Christensen. *The Innovator's Dilemma: When New Technologies Cause Great Firms to Fail*. Harvard Business Review Press, 1997.

[8] Stefan Wuchty, Benjamin F. Jones, and Brian Uzzi. The Increasing Dominance of Teams in Production of Knowledge. *Science*, 316(5827):1036–1039, May 2007.

[9] Michael A. Cusumano, Yiorgos Mylonadis, and Richard S. Rosenbloom. Strategic Maneuvering and Mass-Market Dynamics: The Triumph of VHS over Beta. *Business History Review*, 66(1):51–94, 1992.

[10] Hayagreeva Rao, Philippe Monin, and Rodolphe Durand. Institutional Change in Toque Ville: Nouvelle Cuisine as an Identity Movement in French Gastronomy. *American Journal of Sociology*, 108(4):795–843, January 2003.

[11] Paul J. DiMaggio and Walter W. Powell. The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields. *American Sociological Review*, 48(2):147–160, 1983.

[12] Robert K. Merton. *The Sociology of Science: Theoretical and Empirical Investigations*. University of Chicago Press, 1973.

[13] Michael L. Tushman and Philip Anderson. Technological Discontinuities and Organizational Environments. *Administrative Science Quarterly*, 31(3):439–465, 1986.

[14] Richard J. Rosen. Research and Development with Asymmetric Firm Sizes. *The RAND Journal of Economics*, 22(3):411–429, 1991.

[15] Joshua M. Nicholson and John P. A. Ioannidis. Conform and be funded. *Nature*, 492(7427):34–36, December 2012.

[16] Wesley M. Cohen. Fifty Years of Empirical Studies of Innovative Activity and Performance. In Bronwyn H. Hall and Nathan Rosenberg, editors, *Handbook of the Economics of Innovation*, volume 1, pages 129–213. North-Holland, January 2010.

[17] Michael Lewis. *Moneyball: The Art of Winning an Unfair Game*. W. W. Norton & Company, New York, NY, 1st edition edition, March 2004.

[18] Brad Stone. *The Everything Store: Jeff Bezos and the Age of Amazon*. Little, Brown and Company, New York, 1st edition edition, October 2013.

[19] Lee Fleming. Recombinant Uncertainty in Technological Search. *Management Science*, 47(1):117–132, January 2001.

[20] Brian Uzzi, Satyam Mukherjee, Michael Stringer, and Ben Jones. Atypical Combinations and Scientific Impact. *Science*, 342(6157):468–472, October 2013.

[21] Jacob G. Foster, Andrey Rzhetsky, and James A. Evans. Tradition and Innovation in Scientists' Research Strategies. *American Sociological Review*, 80(5):875–908, October 2015.

[22] Eric Van den Steen. A Formal Theory of Strategy. *Management Science*, 63(8):2616–2636, 2017.

[23] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv:1810.04805 [cs]*, May 2019.

[24] Noah Askin and Michael Mauskapf. What Makes Popular Culture Popular? Product Features and Optimal Differentiation in Music. *American Sociological Review*, 82(5):910–944, October 2017.

[25] Bas Hofstra, Vivek V. Kulkarni, Sebastian Munoz-Najar Galvez, Bryan He, Dan Jurafsky, and Daniel A. McFarland. The Diversity–Innovation Paradox in Science. *Proceedings of the National Academy of Sciences*, April 2020.

[26] Fabio Petroni, Tim Rocktäschel, Sebastian Riedel, Patrick Lewis, Anton Bakhtin, Yuxiang Wu, and Alexander Miller. Language models as knowledge bases? In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 2463–2473, 2019.

[27] R. Rosenfeld. Two decades of statistical language modeling: where do we go from here? *Proceedings of the IEEE*, 88(8):1270–1278, August 2000.

[28] Bryan Kelly, Dimitris Papanikolaou, Amit Seru, and Matt Taddy. Measuring Technological Innovation over the Long Run. Working Paper 25266, National Bureau of Economic Research, February 2020.

[29] Russell J. Funk and Jason Owen-Smith. A Dynamic Network Measure of Technological Change. *Management Science*, 63(3):791–817, March 2016.

[30] Aaron Gerow, Yuening Hu, Jordan Boyd-Graber, David M. Blei, and James A. Evans. Measuring discursive influence across scholarship. *Proceedings of the National Academy of Sciences*, 115(13):3308–3313, March 2018.

[31] Nikhil Garg, Londa Schiebinger, Dan Jurafsky, and James Zou. Word embeddings quantify 100 years of gender and ethnic stereotypes. *Proceedings of the National Academy of Sciences*, 115(16):E3635–E3644, April 2018.

[32] David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022, 2003.

[33] Theresa S. Cho and Donald C. Hambrick. Attention as the Mediator Between Top Management Team Characteristics and Strategic Change: The Case of Airline Deregulation. *Organization Science*, 17(4):453–469, August 2006.

[34] Rui Shen, Yi Tang, and Guoli Chen. When the role fits: How firm status differentials affect corporate takeovers. *Strategic Management Journal*, 35(13):2012–2030, 2014.

[35] Pierre J. Richard, Timothy M. Devinney, George S. Yip, and Gerry Johnson. Measuring Organizational Performance: Towards Methodological Best Practice. *Journal of Management*, 35(3):718–804, June 2009.

[36] Bronwyn H. Hall, Adam Jaffe, and Manuel Trajtenberg. Market Value and Patent Citations. *The RAND Journal of Economics*, 36(1):16–38, 2005.

[37] MacGregor Knox and Williamson Murray. *The Dynamics of Military Revolution, 1300–2050*. Cambridge University Press, August 2001.