Measuring Founding Strategy *

Jorge Guzman

Columbia University †

Aishen Li

Columbia University ‡

Abstract

We propose an approach to measure strategy using text-based machine learning. The key insight is that distance in the statements made by companies can be partially indicative of their strategic positioning with respect to each other. We formalize this insight by proposing a new measure of strategic positioning—the strategy score—and defining the assumptions and conditions under which we can estimate it empirically. We then implement this approach to score the strategic positioning of a large sample of startups in Crunchbase in relation to contemporaneous public companies. Startups with a higher founding strategy score have higher equity outcomes, reside in locations with more venture capital, and receive a higher amount of financing in seed financing events. One implication of this result is that founding strategic positioning is important for startup performance.

Keywords: *Entrepreneurship , Strategy , Strategic Positioning , Machine Learning , Text Analysis*

*We thank Mabel Abraham, Natarajan Balasubramanian, Raj Choudhury, Bo Cowgill, Enrico Forti, Kristina McElheran, Daniel Keum, Bruce Kogut, Robert Seamans, and Zhiyuan Yu for helpful comments, as well as participants in the Columbia Data Science Day and the NYU Workshop on Machine Learning and Strategy.

†Jorge Guzman: Uris Hall, 711. Columbia University. Broadway and 116th. New York, NY. 10027. jag2367@columbia.edu.

‡Aishen Li: Columbia University. Data Science Institute. al3836@columbia.edu.

"Competitive strategy is about being different."

Michael E. Porter (1996)

1 Introduction

The field of strategic management is built around the idea that a better strategy predicts higher firm performance.[1] A formidable amount of strategy research therefore seeks to explain what characteristics of firms and environments make for a better strategy, and what choices do managers have to make to achieve it. Yet, in contrast to this monumental work, there appears to be hardly any research proposing how to measure whether a firm has good a strategy.[2] Even within a single school of thought—such as, strategic positioning—researchers have no step-by-step approach to score a strategy and represent whether it is 'better' or 'worse' than other firms. If we cannot measure whether a firm is better positioned than another, and if we cannot show this predicts follow-on perfomance, how do we know positioning actually matters? This obvious question appears to be missing an empirical answer, and its absence is reflected in debates across the field. In entrepreneurial strategy, for example, the recent push for an experimentation view (Kerr et al. 2014, Koning et al. 2019, Gambardella et al. 2018) has led many to propose that positioning might not matter due to the significant uncertainty that startups face and opportunities for adaptation (Reis 2011). Others have instead come to its defense, emphasizing how founding choices even determine the path of experimentation itself (Felin et al. 2019, Gans, Scott & Stern 2019). An ability to measure strategic positioning would shed light into this and many other questions.

In contrast to the absence of an algorithmic approach to measure strategy, human analysts themselves often appear quite capable of assessing and comparing the positioning of firms. They do this, for example, by listening to companies' statements, where companies tend to emphasize their differences and unique value proposition. Consider Southwest Airlines and Delta Airlines. Southwest's slogan is "Low fares. Nothing to hide. That's TransFarency!". This slogan emphasizes

[1]For example, Porter (1985) defines sustainable competitive advantage as the "fundamental basis [to achieve] above average performance in the long run". (p. 11)

[2]After an investigation the only related paper we have found is Nath & Sudharshan (1994), who measure patterns in choices to understand strategy 'coherence'.

low cost and transparency, which would be particularly appealing to cost-sensitive customers tired of extra fees. Delta instead uses the slogan "World's Most Trusted Airline", which does not focus on low cost but instead on trust and global coverage. Trust and global coverage might not be as valuable for the cost-sensitive travelers to which Southwest caters, but will be for other travelers that instead seek to get anywhere reliably and on time, and are willing to pay extra to do so.[3]

In this simple comparison, a strategy analyst can quickly and intuitively pick the differences between these two statements, even for companies in exactly the same industry. These differences in stated statements are not simply differentiation in the product features. The product in this case is ostensibly similar (a flight), and they instead capture the value proposition, be it variety (in route coverage, for Delta) or cost-leadership (Southwest). Even for companies whose competitive advantage is not a unique product, but the ability to deliver a lower price or variety, these features are emphasized in the company's marketing to consumers. More importantly, if the stategy analyst is given a third company—say Spirit Airlines, which has the slogan "Less Money, More Go."—she would also recognize a difference in the perceived 'distance' between this new statement and the prior two. Spirit Airlines appears closer to Southwest, since both focus on the importance of low cost, and would therefore impinge on the differentiation of Southwest more than of Delta.

Now, consider expanding on this idea to all companies and to a richer set of company statements. If an analyst had the marketing materials of all airline carriers, or even all companies in the United States, might she be able to map systematically the differences and distance between the value proposition of one company and all others? Wouldn't a measure of this distance (or, how 'far' is a company from others) monotonically reflect a better (or worse) strategic positioning? Cognizant of the fact that most companies are unrelated to each other, what would be the best way to aggregate all pair-wise distances to focus on the relationship to the closest competitors, who actually shape competitive dynamics?

In this paper, we build on the idea that distance in how companies talk about themselves can be used to assess their strategic positioning, and use machine learning text-as-data algorithms and

[3]In fact, Southwest has a relatively low rate of on-time arrivals.

historical records to develop a novel approach to measure startup founding strategy. Our key insight is that while it might be almost impossible to observe the value proposition that customers find in the product (or service) offered by a new startup, it is possible, under reasonable assumptions, to see what *the startup itself* believes its value proposition to be. In fact, startups advertise constantly their value proposition through marketing materials in an effort to engage potential customers with their product or service. Virtually all growth-oriented startups founded after a certain date state this value proposition early-on in one very concrete marketing channel: their website. If the value proposition stated online by startups is close enough to their true value proposition, and if we can quantify how differentially positioned is this value proposition from the value proposition of other companies, then we can measure a startup's strategic positioning.

We begin by formalizing this idea in a simple economic model that relates positioning to profits and strategy. Companies gain oligopolistic profits when they can offer something unique to customers that is different from other value propositions in the market, which begets market power. Market power, however, is hard to observe. Instead, we define a new measure, the *Strategy Score*, as any measure that monotonically predicts higher expected market power. This is the definition of competitive advantage—the organization of a firm to expect future positive profit streams. We then take advantage of the idea that companies state their own value propositions to develop a methodology that uses off-the-shelf machine learning tools to estimate the similarity between the statements made by startups and those made by incumbents.[4] The resulting estimate is a pair-wise number, indicating how close (or far) the stated value proposition is between each startup and incumbent pair. We use a simple heuristic to aggregate these pair-wise scores into a firm-level score. Building from studies in industrial organization showing that competitive dynamics are mostly shaped by the first three to five competitors (Bresnahan & Reiss 1991, Igami & Uetake 2019), we define the average distance form the closest five incumbents as the strategy score. If measured with the statements at firm founding, it is the *Founding Strategy Score*, the key measure of our paper.

We implement this approach for a large population of startups from Crunchbase and all U.S.

[4] We use a term frequency inverse document frequency (tf-idf) algorithm to create a weighted measure of all words in each of the startup and incumbent statements, and then cosine similarity to create a scalar measure of similarity.

firms publicly listed at the time of founding. Using the Wayback Machine from the Internet Archive, we download the 'about us' (or equivalent) page on each of the startup's founding websites as it was written close to the time of founding, and compare this to the business purpose section of all 10-K annual reports filed by public companies in the year the startup was founded. We find the average value of similarity between a random startup-incumbent pair is quite low. Though our measure is defined between zero and one, the mean is only 0.012 with a standard deviation of 0.03. However, this value increases to 0.12 at the 99th percentile and 0.31 at the 99.9th percentile—a few incumbents do relate closely to startups. We define distance as the arithmetic inverse (one minus) similarity, and aggregate this into a founding strategy score using our heuristic of the average distance to the five closest incumbents. We find that our estimated score has ample variation, going from 0.66 to 0.86 from the 10th to 90th percentile. This score is highly correlated across alternative aggregating functions, including changing the number of close incumbents considered or using market value as weights when averaging the distances. Geographically, firms with a higher founding strategy score are more likely to be located in geographies with higher venture capital. Then, we proceed to assess the ability of our score to predict firm performance. In models controlling for founding year and city fixed-effects, founding strategy predicts equity growth events (IPO or acquisition). These effects appear to be particularly driven by an increase in the probability of exit at the high end of our measure, rather than differences in actual value upon exit. Our measure also predicts financing performance at the early stages. Even in the first seed round of each company, founding strategy score predicts higher fundraising after controlling for a series of detailed fixed-effects.

This paper contributes conceptually by providing a formal way to score strategy that builds from the basic tenets of strategic positioning. The new measure we propose, the strategy score, is different than prior attempts at formalizing strategy in that we focus specifically in the caliber of the strategy (the predictability of follow-on profits), while prior work has instead mostly sought to define the nature of what is strategy (Van den Steen 2016, Porter 1996). There are many applications for a score of strategic differentiation in research, and we hope that our results here provide useful guidance to researchers that simply seek to understand better or worse strategies without so much focusing on how, specifically, each of the companies differ.

We consider the most important contribution of our paper as methodological. We move from a conceptual idea, strategic positioning, to create a measure that can be estimated using off-the-shelf machine learning tools and common datasets, and we validate our measure in a specific setting in which the results are of interest: entrepreneurship. Even though there is a boom in the use of text-as-data in social science (Gentzkow et al. 2019), sometimes applied to understanding differentiation (Hoberg & Phillips 2010, 2016, Menon et al. 2018), our implementation focusing specifically on strategic positioning is distinct from this prior work, and these differences are critical for its use in strategic management. To support follow-on research in building from our methodology, we have publicly released all the code used to build our founding strategy score, and the estimated word-specific weights in similarity estimates.[5] We believe we are just starting down an important path, and expect follow-on work will build on our approach and improve on it.

Finally, our results also provide empirical insight on the broader discussions of whether and how positioning matters in entrepreneurial strategy. In particular, a recent literature on experimentation has emphasized that there is significant external uncertainty in firm strategy (Kerr et al. 2014, Koning et al. 2019). A common concern is that, at its extreme, startup performance might simply consitute a random pick from an skewed distribution (and hence maximized by simply doing more experiments),[6] or that due to high uncertainty and avenues for experimentation positioning itself might not matter (Reis 2011). Instead, consistent with Gans, Scott & Stern (2019), our results show that strategic positioning at founding does matter. Even in the uncertainty of early startup strategy, there are meaningful differences in the way firms think about themselves and state this proposition, which in turn predict performance. Though a positioning and an experimentation approach are not by definition contradictory (e.g., a firm's intent to learn and experiment could be part of its founding strategy), our results do add nuance to the understanding of how successful

[5] All code is available at https://github.com/SAL116/strategyscore

[6] A longer literature building on Gibrat's Law has emphasized that most growth dynamics appear consistent with a pattern of random proportional growth (see Sutton (1997) for a review).

startups develop, and emphasize the importance of founding choices for follow-on performance (Guzman & Stern 2019).

The rest of the paper proceeds as follows. Section 2 reviews the literature. In Section 3, we use an economic model to formalize our methodological approach. Section 4 presents our results. Finally, Section 5 concludes.

2 Theoretical overview: Can we measure startup strategy?

At least since Porter (1980), strategy research has focused on understanding the causes of competitive advantage[7]—i.e., the ability of a firm to consistently attain superior economic performance.[8] Competitive advantage is created by offering something unique to consumers, and organizing the company in a set of interrelated activities that deliver this value proposition. As Bachmann says, in competitive strategy, "It's OK to be different" (Bachmann 2002).

While a substantial portion of strategic management research and teaching has focused on understanding what creates competitive advantage, measuring (or scoring) the degree of competitive advantage itself has remained difficult. Perhaps this is in part expected. Strategy is a field that borrows from multiple disciplines (economics, sociology, psychology) and hence does not directly conform to the quantitative methods of analysis developed in each one. However, the definition of strategy—at least the classical 'positioning' definition of Porter—is very straightforward and amenable to being quantified: a strategy is how a firm offers something different that allows it to gain profitability. A talented analyst well-versed in the analysis and theory of strategy should be able to choose 'better' and 'worse' strategies. If strategies can be ranked, they can be, in principle, scored.

Even though measuring strategic positioning is difficult for all firms, it would appear even more challenging for entrepreneurial strategies, particularly around the time of founding. One dominant view is best summarized by Hathaway and Litan, "The problem is that it is very difficult,

[7] As Barney (2002) states "Before *Competitive Strategy* there was little consensus about the objective of strategic management research and practice... [The book] marked a fundamental change in the field of strategic management as we have come to know it."

[8] In turn, superior economic performance was defined as a rate of return "in excess of the return on government securities adjusted by the risk of capital loss" (Porter 1980).

if not impossible, to know at the time of founding whether or not firms are likely to survive and/or grow. This is true even with venture-capital backed firms" (Hathaway & Litan 2014). On the other hand, recent research has highlighted at least some systematic and observable differences across firms at founding. Anthony et al. (2016), for example, present compelling qualitative evidence in the early synthetizer industry showing how different messages are used by different startups to create different value propositions for potential customers. As well, a series of recent papers using systematic business registrations shows how simple founding choices such as holding intellectual property or the choice of jurisdiction can explain a large portion of the variance in growth outcomes across firms (see Guzman & Stern (2015, 2019) and related work). How could one build from these insights to score startup founding strategy in a systematic way and study its impact on firm performance? This is the exercise to which we now turn.

3 Measuring startup strategy: a text-based approach

Our measurement approach builds on the idea that written statements by firms partially reflect their strategic positioning. The idea of using firm statements to understand competitive dynamics was pioneered in economics by Hoberg and Phillips, who use 10-K statements to develop new measures of industry categorization based on public firm similarity (Hoberg & Phillips 2016), product market fluidity (Hoberg et al. 2014), and industry choice (Hoberg & Phillips 2017). Their work provides strong and compelling evidence that public text data created endogenously by firms contains significant market information. Our approach is distinct from their work in its focus on strategy rather than industry,[9] in proposing a different aggregating function based on industrial organization that focuses on the closest competitors, in its implementation on entrepreneurial firms—including being able to download founding websites and showing they contain strategic information—, and in using our estimates to learn whether founding stategy predicts startup performance. In this section, we first describe the theoretical setup and define a new concept, the strategy score. Then,

[9]The difference is not only conceptual, but instead changes the way to use text-as-data to focus on strategy. For example, while Hoberg and Phillips use a pre-defined corpus, we let the corpus develop endogenously from the sample of firms. This allows us to focus on non-product words that are nontheless central to strategic differentiation, such statements about low cost, or variety in the companies' statements.

we explain how to use firm statements to assess overlap in the value propositions of firms, and how to translate this to a measure of distance. Finally, we explain how to aggregate measures of distance into a specific measure of the strategy score, allowing us to estimate the founding strategy of startups.

Setup

Consider a startup i with some value proposition that depends on their product (or service), the startup's cost structure and price, and the way in which the startup delivers this product. At the time it is founded, there are J incumbents, indexed by j, already present in the market. The value proposition of each startup and incumbent is different, but may be related. The consumer has an elasticity of substitution across value propositions (not products) ε_{ij} that reflects how its preferences vary between the startup and the incumbent. These elasticities of substitution are aggregated through some function g to estimate the available market space for the startup to attain market power M_i

$$M_i = g(\varepsilon_{i1}, ..., \varepsilon_{iJ}) \tag{1}$$

The firm realizes a performance outcome (such as profit) based on this market power and the underlying demand for its product or service, D_i, and a random term μ_i.

$$\Pi_i = h(M_i, D_i)\mu_i \tag{2}$$

We assume this market power comes from a good strategy and, more specifically, a strong positioning in the market that relates to the ability to have high *expected* profits.

Definition 1 (Strategy score)**.** *For any company, a measure of the caliber of their strategy, their 'strategy score', $S_i > 0$ is a scalar measure that can be positive and monotonically translated to higher market power through some positively increasing function ζ.*

$$M_i = \zeta(S_i) \ , \ \frac{\partial \zeta}{\partial S_i} > 0 \ \forall S_i \tag{3}$$

The purpose of our approach is to develop a data-driven way to estimate an empirical equivalent to S_i.

Measuring similarity through firm statements

Estimating elasticities of substitution, however, has proven substantially difficult. A challenge which is heightened in our setting since we are focused on differences in the substitutability across value propositions (rather than products and their features), and the observability and codification of the key characteristics of these propositions.

Our approach solves this challenge by noting three important insights. First, while it is virtually impossible to observe the value a consumer sees in a product, it is much easier to observe what the firm *believes* its value proposition to be. In fact, firms constantly make statements about what their value proposition is with the goal of explaining to consumers (or to some representative set of them) why the product is valuable, and why it should be purchased. A high level of similarity between the statements of two firms by and large indicates a level of overlap in value propositions. Second, measuring similarity is not merely a theoretical idea: there are standard text-analysis algorithms that allow us to quantify statements by companies and assess the similarity between different texts, effectively allowing us to create a measure of similarity in the *stated* value proposition of firms in the market. These algorithms have already been shown empirically to work well for estimating relatedness across firms in 10-K statements (Hoberg & Phillips 2016), and patents (Bowen III et al. 2018), so applying them to startups is a natural next step. Third, observing at least some of these statements is possible. The IT revolution has created observable records that allow us to retrieve statements by startups and incumbents around the time of startup founding. The most obvious sources of company statements are annual reports and company websites.[10]

Building on these insights, we introduce a new measure of market relatedness, *Similarity*, represented as σ_{ij}. Given a startup and an incumbent statements s_i and s_j (one each) explaining their

[10]Of course, firms say many things in their annual reports and websites. The goal of our statistical approach will be understanding how to use this text to measure differences in value proposition (and its distance to other firms), while giving less weight to other aspects that are unrelated and might therefore not predict performance. A key choice we made towards this goal, for example, is using only nouns and their counts and not include any measures of sentence structure, sentiment, and writing approach in our algorithm.

main value proposition, there exists some function h defined between 0 and 1 that can measure a pair-wise similarlity between these two statements as

$$\sigma_{ij} = h(s_i, s_j), \sigma_{ij} \in [0,1]$$

Companies with a value of similarity equal to 1 have completely equivalent statements, while companies with a similarity of 0 have no relationship to each other. Companies with partial similarity are in between. While there exist many potential functions to estimate σ_{ij}, we simply follow established best practices in text comparison algorithms and define h as the cosine similarity (normalized dot product) between the word counts in each of these statements weighted by the term frequency-inverse document frequency (tf-idf) algorithm.[11]

In our implementation, we begin with a collection of all websites for firms founded in a specific year and the Section 1 (the business decription) of all the public 10-K reports issued in that year. We then keep only nouns from these company statements,[12] removing all words that appear in more than 60-percent of documents, de-steming them, and tokenizing them using standard libraries.[13] Next, we use the incidence of each word across the full set of statements, and estimate a weight for each word according to the inverse of its document frequency. Words that are less common across documents are more informative, and therefore receive high weights, while common words receive low weights. We then transform each statement into a word vector representing each possible word, and with a value equal to the number of times the word is mention in the document times its inverse frequency weight. Finally, we estimate similarity through the cosine similarity between the two weighted vectors. This estimate, defined between 0 and 1, is the tf-idf score. Our code

[11]Another option could be to use Latent Dirichlet Allocation (LDA), a popular machine learning topic model to classify documents. In this probabilistic model, topics are not strongly epistemologically defined, rather identified from term co-occurrence. In practice, LDA is best for topics such as "heart disease", "property insurance", and "modern art" etc. In our study, we want to differentiate value propositions in the most granular level, and simply utilizing the tf-idf method can achieve our goal in the best way.

[12]We only used nouns in this approach since we want to minimize the sentiment and style effect, which mostly carried by adjectives or adverbs. Table 1(A) shows summary statistics for nouns distribution over our 10-K text data set.

[13]We use the following Python libraries. To filter out stop words, we use the `stop_words.get_stop_words` function; to tokenize them, we use `nltk.tokenize.RegexpTokenizer` and simply separate words; to de-stem we use the function `nltk.stem.WordNetLemmatizer`.

implementing this approach in Python is available through Github.[14]

We consider this estimated similarity a monotonic (though noisy) representation of the elasticity of substitution between value propositions, and state the following proposition without proof.

Proposition 1. *Companies with a higher level of similarity in their statements are also more likely to have a higher elasticity of substitution between their value propositions.*

$$\frac{\partial E[\varepsilon_{ij}|\sigma_{ij}]}{\partial \sigma_{ij}} > 0 \tag{4}$$

From similarity to founding strategy

The next step is to aggregate pair-wise similarity between all startups i and incumbents j into an firm-level strategy score. Since our goal is to capture how different the value proposition is between two companies, rather than how similar, we begin by defining distance, δ_{ij}, by algebraically inverting similarity. Distance is therefore a value between 0 and 1, where 0 means two companies are exactly the same and 1 means they are completely different.

$$\delta_{ij} = 1 - \sigma_{ij} \tag{5}$$

Next, we aggregate distance across all incumbents to get an empirical measure of the strategy score at founding (i.e., an $\hat{S}_i$). The mean or median are not good ways to aggregate measures of competitive overlap because most companies are unrelated to each other. Empirical studies in industrial organization highlight how the dynamics of competition are influenced by a small number of competitors and how, as this number increases, the ability of firms to charge margins quickly decreases approximating fully competitive economy (in strategy parlance, they lose their competitive advantage). In this paper, we follow a simple heuristic and use the classic finding of Bresnahan & Reiss (1991) showing that markets become competitive after the first three to five competitors.[15] While this heuristic is admittedly ad-hoc and imperfect, it allows a simple tractable

[14] All code and approach is available at `https://github.com/SAL116/strategyscore`.

[15] In more recent, Igami & Uetake (2019) also study the impact of competition on incentives to innovative and find

approach that is applicable across many firms.[16]

The empirical strategy score can then be estimated by:

$$\hat{S}_i = \frac{1}{5} \sum_{j \in J_i^5} \delta_{ij} \; , \; J_i^5 = \{5\, closest\, incumbents\} \tag{6}$$

While $\hat{S}_i$ can in principle represent the strategy score at any point in time, we are estimating the value at startup founding—the *Founding Strategy Score*. Which is an imperfect measure of the strategic positioning of a startup at the time it is founded, and translates to market power due to a unique positioning in the market vis-á-vis incumbents.

4 Data: Crunchbase, the Internet Archive, and 10-K statements

To implement our approach, we build a dataset that includes a comprehensive list of startups from Crunchbase, their historical websites at the time of founding retrieved through the Internet Archive, and the annual 10-K statements of firms from the U.S. Securities and Exchange Commission EDGAR system.[17] We describe each dataset in turn.

Startup data. We downloaded all companies available in Crunchbase founded between 1997 and 2018 that have a website and have received some form of financing. Crunchbase is a popular crowd-sourced data platform tracking a large number of innovation-based startup companies, their website, founding date, funding rounds, and equity outcomes over the last two decades. It is one of the main databases used in entrepreneurship and strategy research, and performs particularly well in covering firms in what might be considered the 'startup economy' (Dalle et al. (2017) provide an overall assessment and examples of the use of Crunchbase in managerial and economic research).

For each company, we downloaded in April of 2019 the company name, the founding date, the city, the date and amount of each seed financing round, whether the company achieved an equity

that incentives to innovative drop quickly, stabilizing after five competitors.

[16]In the empirical section, we show that this is not sensitive to the specific number of five competitors, or to whether the averaging approach weights each incumbent by its market value during the year of founding.

[17]EDGAR stands for Electronic Data Gathering, Analysis, and Retrieval system.

event (IPO or an acquisition), the market valuation of the company at the exit event, the timing of the exit event, and the top level Crunchbase category for this firm. Categories are a sort of industry categorization created by Crunchbase, but made especially to better account for the heterogeneity across startups. We refer to it hereafter as ‘industry’. The total number of firms was 36,462, of which we were able to download a website for 12,103 firms, distributed across 498 industries, with the top ones being e-commerce (839 firms), internet (778 firms), health care (72 firms), education (540 firms), and SaaS (369 firms).

Website history data. We used the Wayback Machine, an online platform offered by the Internet Archive (archive.org), to download the initial website of each startup around the time of founding. The Wayback Machine provides access to a digital library containing over 330 billion web-page snapshots occurring in history. These snapshots are taken at least a few times a year for all unique domain names on the internet. We developed a web-scraping technology to automatically query the Wayback Machine for the earliest version of the webpage in the year after the year of founding in Crunchbase. We visited this historical version of the website and searched for sub-pages whose link contains any of the keywords ”About”, ”About us”, ”Company”, ”Products”, ”Overview”, ”Services”, ”Features”, ”Solutions”, ”What we do”, and ”Who we are”, in that order. We downloaded the first matching page and performed a cleaning procedure to keep only the written text (no HTML). If no informative sub-pages were found, we use their homepage, and if one of the sub-pages had an error or was too small (less than 1,000 characters of text), we simply ignored it and tried the next possible option. After some other quality checks and small considerations, we successfully downloaded the company’s own description of themselves in their founding website for 12,103 startups.

Incumbent information. Finally, to observe the incumbents’ business proposition, we took advantage of the statements written by publicly listed companies in their 10-K annual filings. We used the Item 1 in each annual report (the company’s business description) for the annual report filed in the year the startup was founded, and compared the startup text with the business description of the incumbent to build a pairwise similarity measurement. Our resulting dataset contains all 10-K filings since 1997, the earliest year in which they are available online. We have a total of 148,235 annual reports, averaging 6,738 per year.

Table 1: 10-K and Startup Summary Statistics

Statistic	Mean	St. Dev.	N
Panel A: 10-K Public Firm Statements			
Words in 10-K Statement	9,708.99	10,268.50	148,235
Panel B: Startup Data			
Statistic	Mean	St. Dev.	N
Equity Growth (IPO or Acquisition)	0.17	0.38	12,103
Acquired	0.15	0.36	12,103
Acquisition Price	$328,729,522.00	$546,141,773.00	372
IPO	0.02	0.15	12,103
Valuation At IPO	$750,697,798.00 7	$26,811,423.00	102
Panel C: Seed Funding Data			
Statistic	Mean	St. Dev.	N
Seed Funding	$1,242,632.00	$16,383,045.00	9,117
First Seed Funding	$1,280,028.00	$9,690,472.00	6,651

Note: The data is built from annual 10-K statements of all public firms and Crunchbase startups. Startup Data is a dataset including one observation per startup. Seed Funding Data is a data on all seed funding events in Crunchbase for those startups that receive seed financing.

Summary statistics

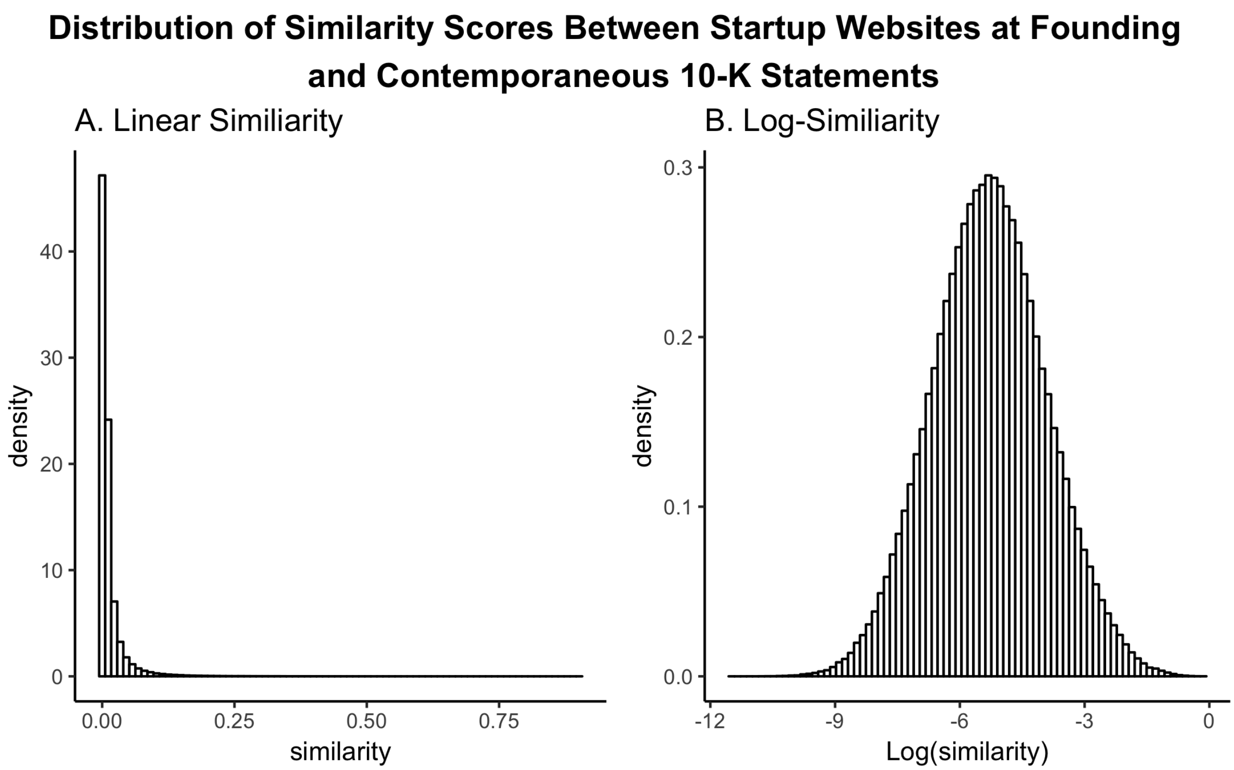
Table 1 presents summary statistics of our data. 17% of our companies achieve an equity growth outcome, with 15% of this being acquisitions and 2% IPOs. The average acquisition price is $328 million dollars, and the average valuation at IPO is $751 million.[18] Both measures are significantly skewed. On average, the firms raise $1.2 million dollars in their seed funding rounds.

[18] As with all acquisition databases, the acquistion value is not known for all acquisitions in Crunchbase, but instead only a portion.

5 Results

We now proceed to estimate founding strategy and its relationship to startup performance. We begin by reporting estimates of similarity across all pairs of firms and how these build to a founding strategy score per company. Then, we study the ways in which the founding strategy score is associated with regional differences, startup equity performance, and financing outcomes.

Figure 1



Notes: The estimated similarity follows a log-normal distribution due to the design of the standard tf-idf implementation. For a word ω_{ij} *that occurs* j *times in document* j*, the weight is* $\omega_{ij} = tf_{ij} + log(\frac{N}{df_i})$

Estimates of similarity and founding strategy

Figure 1 plots the overall distribution of the text similarity score of 67 million startup-incumbent pairs. Summary statistics are reported in Panel A of Table 2. The left panel of Figure 1 is the distribution of similarity. It is easy to see that estimated similarity is very skewed, and has very small values for any pair of startup and incumbent. The mean value of similarity is 0.013, and the median is 0.005. Consistent with high skewness, the standard deviation is twice the mean, at 0.026. The right panel of Figure 1 plots the log-distribution. The log-similarity score appears normally distributed, suggesting that the distribution of our text-based measure is log-normal. Of note, this skewed distribution is a natural outcome of the tf-idf text similarity algorithm, but one of which we take advantage to identify the few companies that are close to each other versus the rest.[19]

Table 2: Estimated Score Summary Statistics

Panel A: Similarity Score Summary Satistics

Statistic	Mean	St. Dev.	Median
Similarity (67 million pairs)	0.0129	0.0264	0.0050

Panel B: Founding Strategy Score Summary Statistics

Statistic	Mean	St. Dev.	N
Founding Strategy Score (7 closest)	0.78	0.09	12,103
Founding Strategy Score (6 closest)	0.77	0.09	12,103
Founding Strategy Score (5 closest)	0.76	0.09	12,103
Founding Strategy Score (4 closest)	0.75	0.10	12,103
Founding Strategy Score (3 closest)	0.74	0.10	12,103
Weighted Strategy Score (5 closest)	0.76	0.10	12,103

Note: Founding Strategy Score is estimated as the mean text distance to the n-closest contemporaneous public incumbents. Weighted Strategy Score is the same but the mean is weighted by contemporaneous market value. Text distance is one minus the cosine similarity of words on the startup website and the incumbent 10-K statement.

Strategic distance is the arithmetic inverse of similarity (one minus similarity), representing the level of market differentiation between a given startup and incumbent. We estimate *Founding Strategy Score*, by averaging the *Nth* closest competitors according to our estimate of strategic distance. Panel B of Table 2 reports the summary statistics of these scores for values of *N* ranging from 3 to 7, as well as a complementary measure that weights the average based on the mean market capitalization of each incumbent that year. The average estimated founding strategy score is between 0.74 and 0.78, depending on the measure, and a standard deviation of 0.09 or 0.10. The

[19]Specifically, the standard tf-idf algorithm assumes similarity across documents is distributed lognormal in estimating the similarity scores.

Table 3: Correlation of Founding Strategy Scores

	(1)	(2)	(3)	(4)	(5)	(6)
(1) Founding strategy score (7 closest)	1					
(2) Founding strategy score (6 closest)	0.996	1				
(3) Founding strategy score (5 closest)∗	0.990	0.995	1			
(4) Founding strategy score (4 closest)	0.976	0.988	0.994	1		
(5) Founding strategy score (3 closest)	0.950	0.967	0.979	0.994	1	
(6) Weighed by market value (5 closest)	0.909	0.915	0.919	0.917	0.907	1

Note: Founding Strategy Score is estimated as the mean text distance to the n-closest contemporaneous public incumbents. Weighted Strategy Score is the same but the mean is weighted by contemporaneous market value. Text distance is one minus the cosine similarity of words on the startup website and the incumbent 10-K statement.

correlation amongst them, shown in Table 3, is very high, above 0.9 for all pairs of measures. We interpret this as suggesting that any of these measures appears to capture a very similar dynamic of the founding strategy score, and choose to use the 5 closest competitors as a benchmark. However, for robustness, we also present in the appendix all our results using the founding strategy score weighted by market value.[20]

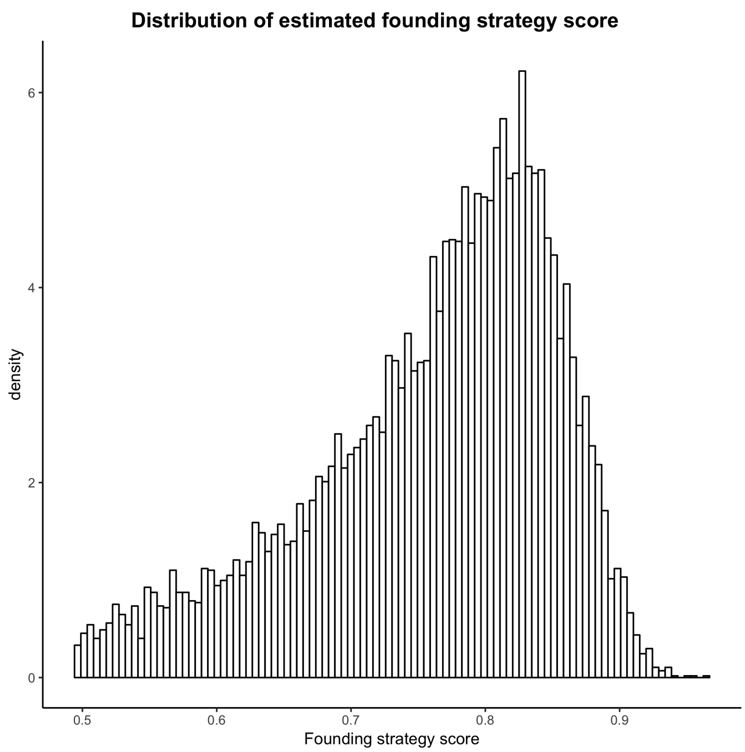
Figure 2 plots the distribution of the estimated strategy score. The distribution is of Weibull shape, which is expected given that this is a maxima estimate of a (log) normal distribution. We drop all firms that have a founding strategy score two standard deviations below the mean.[21]

In Figure 3, we also consider the geographic distribution of our measure by plotting, for all US cities, the mean and median founding score and log total VC financing for startups in our sample. The fitted line is weighted by the number of companies funded in each location. The slope of the fitted line for Panel A is 49.0, suggesting that cities with a mean strategy score that is 10 percentage points higher are associated with 4.9 times more venture capital. Similarly, comparing cities in the 75th percentile of mean strategy score to cities in the 25th percentile, there is about a 96% higher level of total venture capital financing invested. Panel B shows that using the median founding

[20] This choice was based on reading of the IO literature, which suggests that the quantity of competitors is more important than their size in competitive dynamics.

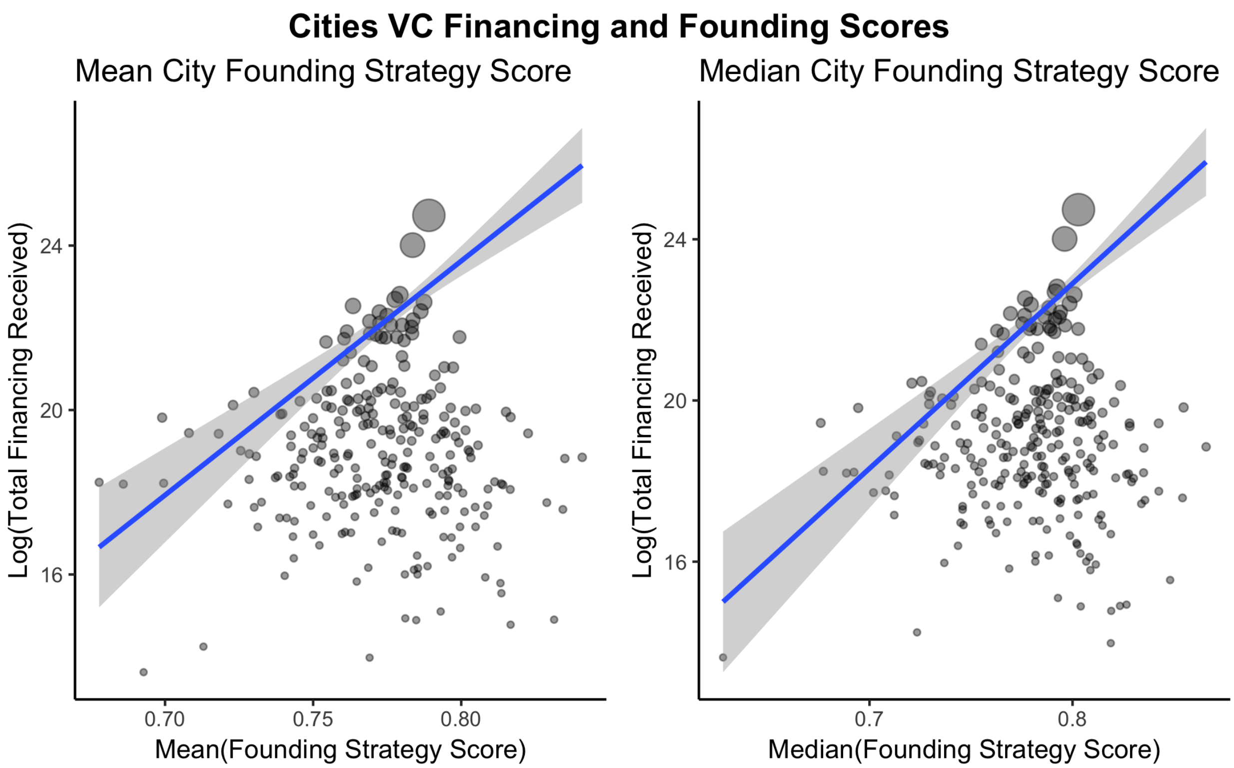
[21] We choose to do this because we empriically found that those firms scoring significantly low on their strategy score had a generic website (e.g., a Go Daddy domain page), rather than their business page.

Figure 2



Notes: The estimated founding strategy score is the average of the highest 5 distance values for each firm. The distribution approximates a Weibull distribution, which is a consistent Extreme Value Type-I distribution of normal estimates.

Figure 3



Notes: Plots the average estimated founding strategy score for all startups in the Crunchbase data founded in a specific city (municipality), versus the total amount fundraised in these cities. Linear trend is weighted by city size according to Log(Total Financing Received).

score produces a very similar pattern.

Founding strategy score and firm equity outcomes

Table 4: Founding Strategy Score and Startup Performance

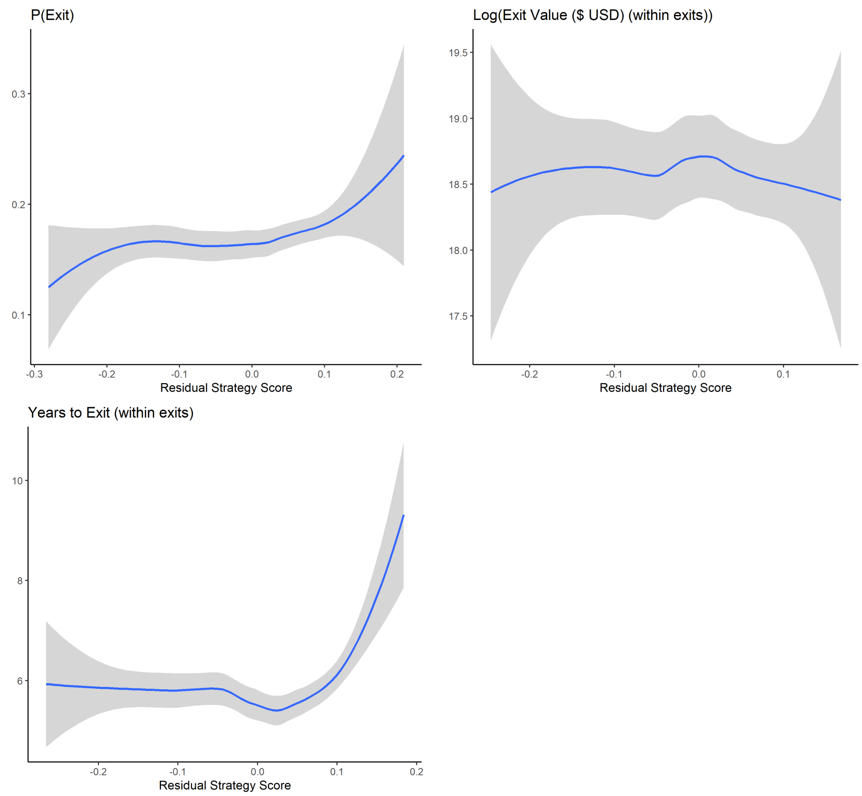
	Dependent variable:				
	Equity Growth (IPO or Acquisition)				
	OLS	OLS	OLS	OLS	Logit
	(1)	(2)	(3)	(4)	(5)
Founding strategy score	-0.073^{**}	0.074^{**}	0.048^{*}	0.066^{**}	0.580^{**}
	(0.037)	(0.033)	(0.028)	(0.030)	(0.282)
Founding Year F.E.	No	Yes	Yes	Yes	Yes
City F.E.	No	No	Yes	Yes	No
Industry F.E.	No	No	No	Yes	No
Observations	12,103	12,103	12,103	12,103	12,103
R^2	0.0003	0.097	0.191	0.229	
Log Likelihood					−4,919.094

Note: **p<0.1;* ***p<0.05;* ****p<0.01*
City F.E. reflects individual fixed-effects for each municipality. Industry F.E. uses the top-level categorization provided by Crunchbase. Standard errors double clustered by city and founding year for OLS regressions, and clustered only by founding year for logit model.

We now consider whether founding strategy score predicts firm outcomes. In Table 4, we report the relationship of founding strategy score on the probability that a firm achieves equity growth. Standard errors are double clustered by year and city to account for potential correlations across firms on either of these dimensions.

Column (1) reports the coefficient of a simple OLS regression with no controls to assess the overall correlation between the two variables. The coefficient is negative and significant. This is not intuitive. Column (2) resolves this conundrum by adding founding year fixed-effects to account for differences in the age of firms and the business cycle, and the coefficient turns positive. Column (3) includes city fixed-effects to account for the possibility of differences in location that drive differences in performance in ways unrelated to actual strategy. The coefficient stays positive and significant, with a value of 0.048 and significant at the 10% level. Column (4) adds industry

Figure 4: Smoothed Relationship of Founding Strategy Score and Firm Performance



Notes: Smoothing estimator is locally weighted smoothing (loess). Residual Novelty is the residual from an OLS regression of Founding Strategy Score on founding year fixed-effects to control for differences in the business cycle that could correlate to outcomes and underlying strategy scores.

fixed-effects, resulting in a more precise coefficient with a value of 0.066 significant at the 5% level. Finally, Column (5) repeats the specification as a logit model, showing the same effects.

Given a mean of *Equity Growth* of 0.17, these estimates are economically meaningful. Taking Column (4) as the preferred estimate, the results suggest that going from 1 standard deviation below the mean to one standard deviation above the mean in the strategy score increases the probability of achieving equity growth by about 10%. Column (5) shows a similar effect for a logit specification that includes founding year and city fixed-effects. The marginal effect of moving two standard deviations is 13%.

In Figure 4 we then consider the heterogeneity across some of these outcomes by plotting smoothed fitted models against our measure. To avoid business cycle effects and other cohort effects, we report the residual of founding strategy after regressing it on founding year fixed-effects. This estimate therefore reflects differences in the estimated strategy score around the average value for firms founded each year. Consistent with a measure of differentiation, we observe that the probability of exit increases monotonically across the novelty distribution, with a sharper upswing towards the top of our measure. However, we do not observe a change in the equity value upon exit (either acquisition price or the valuation at IPO). Interestingly, we also observe that firms that score higher in our measure also take longer to exit in general. [22]

Together, this evidence suggests that our measure predicts startup performance in a way that would have been predicted by the positioning literature, and, therefore, that it appears to be able to measure startup founding strategy.

Founding strategy score and financing

Next, we consider the extent to which our measure also predicts the financing received by startups. Prior work has demonstrated a strong relationship between financing events, and both firm quality (Catalini et al. 2019) and intellectual property strategy (Hellmann & Puri 2000). This prior work has particularly investigated the extensive margin of financing: whether or not firms receive venture capital financing, based on pre-VC startup characteristics. Our data is instead best served to study the intensive margin—the amount of financing received.[23] We choose to focus particuarly on an early financing event, seed financing rounds, to assess the extent to which our founding strategy score also predicts variation in the very early levels of financing. Seed financing events represent the early financing rounds received by a startup, often when a startup is still an uncertain concept. Seed financing funds are relatively small in size (mean of $1.2 million dollars) and are made based on the early potential of the firm as perceived by investors. An additional advantage of using these events is that variation in the amount of equity given (which is not observed) is often less of an issue. In contrast to later financing rounds, seed financings tend to be very standardized in equity

[22] Several potential theories can be offered as to why more differentiated firms might take longer to exit. Rather than hypothesizing on any of them here, we instead choose to leave a fuller assessment of the way strategy shapes the time to exit as an important question for future research.

[23] Crunchbase focuses particularly on venture-backed firms, such that those that never intend to receive financing are much less likely to be in it.

given. They are often either convertible debt[24] or the take of equity in the 20% to 30% range. In contrastlater-on rounds have more variance, so that the funding amount might correlate less cleanly with pre-money valuation.[25]

Table 5: Founding Strategy Score and Initial Financing

	Dependent variable:				
	Log(Seed Funding)				
	(1)	(2)	(3)	(4)	(5)
Founding Strategy Score	0.567**	0.370	0.367**	0.512***	0.289*
	(0.229)	(0.235)	(0.169)	(0.143)	(0.149)
Sample	All	All	All	First Event	All
Founding Year F.E.	No	Yes	Yes	Yes	Yes
City F.E.	No	Yes	Yes	Yes	Yes
Seed Funding Year F.E.	No	No	Yes	Yes	Yes
Industry F.E.	No	No	No	No	Yes
Observations	9,117	9,117	9,117	6,651	9,117
R^2	0.001	0.182	0.223	0.251	0.283

Note: **p<0.1;* ***p<0.05;* ****p<0.01*
City F.E. reflects individual fixed-effects for each municipality. Industry F.E. uses the top-level categorization provided by Crunchbase. Seed Funding Year F.E. are a fixed-effect on the year the funding is received. Sample "All" indicates all seed funding events, while "First Event" indicates only the first seed funding event recorded for this company. Standard errors double clustered by city and founding year.

In Table 5, we study the relationship between all seed financing events and the startup's founding strategy score. To do so, we report OLS regressions with the dollar value of seed funding (*Log(Seed Funding)*) as the dependent variable and *Founding Strategy Score* as the independent variable. Standard errors are double clustered by founding year and city.

Column (1) shows a positive relationship between our measure and total seed financing. The coefficient is 0.567 and significant. Column (2) includes founding year and city fixed effects, to

[24]Convertible debt contracts are contracts under which the valuation will be decided in the future, based on a future financing round.

[25]Relatedly, Ewens et al. (2019) study the structure of individual venture capital contracts and show that total investor equity share (after accounting for other investor rights), is around 40%. However, this matching is often inefficient due to overreaching investors (who maximize their own value at the expense of total firm value). Using early stages (and avoiding Series A and B) helps avoid some of these issues, even if imperfectly.

better assess the relationship of our measure to financing performance within the same cohort of firms, in the same location. Our preferred specification is Column (3), where we include founding year fixed effects, city fixed effects, and fixed effects for the year in which the seed funding event is occuring (*Seed Funding Year F.E.*). Including these fixed-effects allows us to control for any differences in the time that startups take to raise the financing, and therefore their speed. The coefficient can therefore be interpreted as a cohort of startups, born in the same year and location, and getting a seed round around the same time. Our coefficient is positive and significant, with a value of 0.367. This is a meaningful effect for firms: moving from the 10th to the 90th percentile in our founding strategy score would imply an increase of about 10% in the amount of financing received by these startups.

We then consider only the first seed financing event in Column (4). Our coefficient increases to 0.51. Finally, in Column (5) we replicate Column (3) but also include industry fixed-effects to control for differences across the financing dynamics of different industries. Our effects are slightly lower but very similar.

Together these estimates evidence the importance of founding strategy for initial investment. Given the natural impresicion of our measure in actually estimating strategy, we see this estimate as likely a lower bound. We expect true strategy to predict variation in actual financing events much more than the effects we have measured here.

6 Conclusion

Strategy researchers have always taken a normative view on their field. There are better and worse strategies, and the goal of research is to understand what these are and how to help managers develop better ones. The most prevalent view is that a better strategy is that which creates a more differentiated company, allowing it to consistently charge oligopolistic profits. We proposed a novel approach to measure startup founding strategy in a systematic way, and implemented it for a large sample of growth-oriented startups in the United States. Our method takes advantage of the fact that firms themselves state what makes them unique, and that this is stated by most companies close to founding on their website. We estimate the similarity of all startups to contemporaneous publicly listed incumbents, and propose a simple algorithm to aggregate this into a *Founding Strategy Score* for each startup. We show that machine learning and natural language processon tools can be used to imperfectly estimate of this founding strategy score. Using this measure, we find firms with higher founding strategy scores are located in areas that receive more financing, and are more likely to achieve high growth outcomes and receive financing themselves, even conditional on these location and cohort effects.

A conservative view of our work would be to take it almost as a proof of concept. Our goal was to show that strategy can be measured quantitatively using machine learning methods in big data. Measuring strategy appears, to us, as one of the key goals of our field; yet it has proven difficult to achieve. We believe our analysis opens up a conversation on how to use current technology to measure strategy, and we develop quantitative estimates on the impact of founding strategy on performance and the evolution of companies. While we chose to frame our work squarely within the 'positioning' literature, we are well aware that there are other areas of research that instead focus on other drivers of firm performance, such as managerial talent (Bloom et al. 2012) or dynamic capabilities (Eisenhardt & Martin 2000). We expect that other methodological innovations, perhaps building on our approach, will be better at capturing these for young firms.

The correct interpretation of our estimates, however, does require some caution. Importantly, our results do not show, nor do we claim to show, that better strategic positioning actually creates better performance, and it is entirely possible that an omitted variable such as human capital creates both better strategy and (independently) changes performance. While there are obviously potential ways in which one might use our approach to begin to understand the fundamental causes of startup performance, we also believe a different and thoughtful research design would be needed to do this successfully.

On the other hand, taking them only as a predictive result, our estimates probably understate how much strategy does predict performance for several reasons. The first and most obvious one is the imprecise nature of our estimates. We expect others will build on our approach to develop better and more precise scores of founding strategy, which will in turn show a higher explanatory power of strategy on performance. Our sample is also a relatively small set of roughly similar firms (startups in Crunchbase), but it is likely that considering a broader set of firms that includes small businesses would simply make strategy even more relevant. A key challenge in expanding our approach in this direction is that many of these smaller firms serve local markets in non-tradeable goods, so that the competitor set is geographically constrained. This problem was simplified in our data by the assumption that growth startups compete nationally in traded markets, and so most public firms are at risk of being competitors of our startups, even when the startups are small. This assumption would be problematic in non-tradeable local goods. Finally, our empirical approach understated the role of strategy through the use of fixed-effects. These were included in an effort to be conservative and control for potential obvious confounders related to location, cohort, industry, and seed financing year. Yet, location, industry, and the timing of financing, are to an extent chosen strategically by the firm, so the right way to allow this strategic portion to come into the strategy score is unclear. We plan in some future work to try to consider these challenges more clearly.

Theoretically, within entrepreneurial strategy, our results elucidate the existence of limits to the experimentation view of startup strategy (Koning et al. 2019, Kerr et al. 2014, Gambardella et al. 2018). Taken to its extreme, an experimentation view would suggest that founding startup positioning does not matter. The Lean Startup method (Reis 2011), for example, anchors on the assumption that since there is so much founding uncertainty and space for adaptation, a startup can start anywhere in the positioning map, and then simply learn and experiment its way to the top. A critique of this approach, emphasized in prior work on 'rugged landscapes' (Siggelkow 2001), is that firms that build strategy simply by evolving might get stuck in local optimums and miss a global one (Felin et al. 2019). Taking these two views seriously places entrepreneurs in a predicament: founding choices matter, but the extent to which they can choose well might be low due to founding uncertainty (Gans, Stern & Wu 2019). Our analysis shows statistical evidence of the importance of founding strategy, complementing the nascent managerial frameworks developed to better solve this problem (Gans, Scott & Stern 2019).

Ultimately, adjudicating between an experimentation and a positioning view in the design of entrepreneurial strategies, and developing the appropriate tools for managers to develop them, is one of the key frontiers of strategy research. We hope our study has provided much needed empirical evidence to elucidate some of these questions, and a new approach that researchers can build upon to address them.

References

Anthony, C., Nelson, A. J. & Tripsas, M. (2016), '"who are you?… i really wanna know": Product meaning and competitive positioning in the nascent synthesizer industry', *Strategy Science* **1**(3), 163–183.

Bachmann, J. W. (2002), 'Competitive strategy: It's o.k. to be different', *The Academy of Management Executive (1993-2005)* **16**(2), 61–65.
URL: *http://www.jstor.org/stable/4165842*

Barney, J. B. (2002), 'Strategic management: From informed conversation to academic discipline', *The Academy of Management Executive (1993-2005)* **16**(2), 53–57.
URL: *http://www.jstor.org/stable/4165840*

Bloom, N., Sadun, R. & Van Reenen, J. (2012), 'The organization of firms across countries', *The quarterly journal of economics* **127**(4), 1663–1705.

Bowen III, D. E., Frésard, L. & Hoberg, G. (2018), 'Technological disruptiveness and the evolution of ipos and sell-outs', *Swiss Finance Institute Research Paper Series* pp. 19–22.

Bresnahan, T. F. & Reiss, P. C. (1991), 'Entry and competition in concentrated markets', *Journal of political economy* **99**(5), 977–1009.

Catalini, C., Guzman, J. & Stern, S. (2019), 'Passive versus active growth: Evidence from founder choices and entrepreneurial investment'.

Dalle, J.-M., den Besten, M. & Menon, C. (2017), Using crunchbase for economic and managerial research, Working paper, OECD Science, Technology and Industry Working Papers.

Eisenhardt, K. M. & Martin, J. A. (2000), 'Dynamic capabilities: what are they?', *Strategic management journal* **21**(10-11), 1105–1121.

Ewens, M., Gorbenko, A. S. & Korteweg, A. (2019), Venture capital contracts, Technical report, National Bureau of Economic Research.

Felin, T., Gambardella, A., Stern, S. & Zenger, T. (2019), Lean startup revisited, Technical Report 26278, Medium.
URL: *https://medium.com/@teppofelin/lean-startup-revisited-c81fb8719614*

Gambardella, A., Camuffo, A., Cordova, A. & Spina, C. (2018), 'A scientific approach to entrepreneurial decision making: Evidence form a randomized control trial', *Management Science, Forthcoming* .

Gans, J. S., Stern, S. & Wu, J. (2019), 'Foundations of entrepreneurial strategy', *Strategic Management Journal* **40**(5), 736–756.

Gans, J., Scott, E. & Stern, S. (2019), *Entrpreneurial Strategy*, Manuscript.

Gentzkow, M., Kelly, B. & Taddy, M. (2019), 'Text as data', *Journal of Economic Literature* **57**(3), 535–74.

Guzman, J. & Stern, S. (2015), 'Where is silicon valley?', *Science* **347**(6222), 606–609.

Guzman, J. & Stern, S. (2019), 'The state of american entrepreneurship: New estimates of the quantity and quality of entrepreneurship for 32 us states, 1988-2014', *American Economic Journal: Economic Policy* .

Hathaway, I. & Litan, R. (2014), 'Declining business dynamism: It's for real', *Economic Studies at Brookings, Brookings Institution. https://www. brookings. edu/research/declining-business-dynamism-its-for-real* .

Hellmann, T. & Puri, M. (2000), ‘The interaction between product market and financing strategy: The role of venture capital’, *The review of financial studies* **13**(4), 959–984.

Hoberg, G. & Phillips, G. (2010), ‘Product market synergies and competition in mergers and acquisitions: A text-based analysis’, *The Review of Financial Studies* **23**(10), 3773–3811.

Hoberg, G. & Phillips, G. (2016), ‘Text-based network industries and endogenous product differentiation’, *Journal of Political Economy* **124**(5), 1423–1465.

Hoberg, G. & Phillips, G. (2017), ‘Conglomerate industry choice and product language’, *Management Science* **64**(8), 3735–3755.

Hoberg, G., Phillips, G. & Prabhala, N. (2014), ‘Product market threats, payouts, and financial flexibility’, *The Journal of Finance* **69**(1), 293–324.

Igami, M. & Uetake, K. (2019), ‘Mergers, innovation, and entry-exit dynamics: Consolidation of the hard disk drive industry, 1996-2016’, *Available at SSRN 2585840* .

Kerr, W. R., Nanda, R. & Rhodes-Kropf, M. (2014), ‘Entrepreneurship as experimentation’, *Journal of Economic Perspectives* **28**(3), 25–48.

Koning, R., Hasan, S. & Chatterji, A. (2019), Experimentation and startup performance: Evidence from a/b testing, Working Paper 26278, National Bureau of Economic Research.
URL: *http://www.nber.org/papers/w26278*

Menon, A., Choi, J. & Tabakovic, H. (2018), ‘What you say your strategy is and why it matters: natural language processing of unstructured text’, *Acad Manag Proc* .

Nath, D. & Sudharshan, D. (1994), ‘Measuring strategy coherence through patterns of strategic choices’, *Strategic Management Journal* **15**(1), 43–61.

Porter, M. E. (1980), *Competitive strategy: Techniques for analyzing industries and competitors*, Simon and Schuster.

Porter, M. E. (1985), *Competitive Advantage: Creating and Sustaining Superior Performance*, The Free Press.

Porter, M. E. (1996), 'What is strategy?', *Harvard Business Review* **6**(74), 61–78.

Reis, E. (2011), *The Lean Startup: How Today's Entrepreneurs Use Continuous Innovation to Create Radically Successful Businesses*, Currency.

Siggelkow, N. (2001), 'Change in the presence of fit: The rise, the fall, and the renaissance of liz claiborne', *Academy of Management journal* **44**(4), 838–857.

Sutton, J. (1997), 'Gibrat's legacy.', *Journal of Economic Literature* **35**(1), 40–59.

Van den Steen, E. (2016), 'A formal theory of strategy', *Management Science* **63**(8), 2616–2636.

Appendix

Table A1: Founding Strategy Score and Startup Performance

	Dependent variable:				
	Equity Growth (IPO or Acquisition)				
	OLS	OLS	OLS	OLS	Logit
	(1)	(2)	(3)	(4)	(5)
Weighted Strategy Score	-0.063^{*}	0.060^{**}	0.042	0.054^{*}	0.469^{*}
	(0.034)	(0.027)	(0.027)	(0.029)	(0.256)
Founding Year F.E.	No	Yes	Yes	Yes	Yes
City F.E.	No	No	Yes	Yes	No
Industry F.E.	No	No	No	Yes	No
Observations	12,103	12,103	12,103	12,103	12,103
R^2	0.0003	0.097	0.191	0.229	
Log Likelihood					−4,919.532

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$
City F.E. reflects individual fixed-effects for each municipality. Industry F.E. uses the top-level categorization provided by Crunchbase. Standard errors double clustered by city and founding year for OLS regressions, and clustered only by founding year for logit model.

Table A2: Founding Strategy Score and Well-Funded Startups' Performance

	Dependent variable:				
	Equity Growth (IPO or Acquisition)				
	OLS	OLS	OLS	OLS	Logit
	(1)	(2)	(3)	(4)	(5)
Founding strategy score	−0.088	0.093**	0.077	0.113**	0.576*
	(0.055)	(0.046)	(0.052)	(0.058)	(0.333)
Founding Year F.E.	No	Yes	Yes	Yes	Yes
City F.E.	No	No	Yes	Yes	No
Industry F.E.	No	No	No	Yes	No
Observations	6,807	6,807	6,807	6,807	6,807
R^2	0.0004	0.095	0.211	0.263	
Log Likelihood					−3,329.478

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$
Well-Funded Startups are startups with over 1 million USD in total funding. City F.E. reflects individual fixed-effects for each municipality. Industry F.E. uses the top-level categorization provided by Crunchbase. Standard errors double clustered by city and founding year for OLS regressions, and clustered only by founding year for logit model.

Table A3: Founding Strategy Score Across IPO and Acquisition

	Dependent variable:					
	IPO		Acquired		Log(IPO Valuation)	Log(Acquisition Price)
	(1)	(2)	(3)	(4)	(5)	(6)
Weighted Strategy Score	−0.025	0.002	0.066**	0.057**	0.007	−0.062
	(0.017)	(0.018)	(0.029)	(0.027)	(1.958)	(0.290)
Founding Year F.E.	Yes	Yes	Yes	Yes	Yes	Yes
City F.E.	Yes	Yes	Yes	Yes	No	No
Industry F.E.	No	Yes	No	Yes	Yes	Yes
Exit Year F.E.	No	No	No	No	Yes	Yes
Observations	12,103	12,103	12,103	12,103	102	371
R^2	0.175	0.234	0.171	0.208	0.809	0.529

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$
City F.E. reflects individual fixed-effects for each municipality. Industry F.E. uses the top-level categorization provided by Crunchbase. Exit Year F.E. is a different fixed-effect for each year of IPO (in Column (5)) or acquisition (in Column (6)). Including City F.E. in the models of Columns (5) and (6) makes the regression indetermined. Standard errors double clustered by city and founding year.

Table A4: Founding Strategy Score and Initial Financing

	Dependent variable:				
	Log(Seed Funding)				
	(1)	(2)	(3)	(4)	(5)
Weighted Strategy Score	0.412**	0.268	0.317*	0.464**	0.286**
	(0.202)	(0.225)	(0.176)	(0.190)	(0.143)
Sample	All	All	All	First Event	All
Founding Year F.E.	No	Yes	Yes	Yes	Yes
City F.E.	No	Yes	Yes	Yes	Yes
Seed Funding Year F.E.	No	No	Yes	Yes	Yes
Industry F.E.	No	No	No	No	Yes
Observations	9,117	9,117	9,117	6,651	9,117
R^2	0.001	0.182	0.223	0.251	0.283

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

City F.E. reflects individual fixed-effects for each municipality. Industry F.E. uses the top-level categorization provided by Crunchbase. Seed Funding Year F.E. are a fixed-effect on the year the funding is received. Sample "All" indicates all seed funding events, while "First Event" indicates only the first seed funding event recorded for this company. Standard errors double clustered by city and founding year.