ENTREPRENEURSHIP BIAS AND THE MASS MEDIA:

EVIDENCE FROM BIG DATA

INTRODUCTION

Why do individuals choose to pursue an entrepreneurial career? A large body of research has emerged over the last thirty years dedicated to answering this question (Brockhaus, & Horwitz, 1986; Gartner, Shaver, Carter, & Reynolds, 2004; Parker, 2018). While undoubtedly much has been learnt, much about the determinants of entrepreneurial career choice remains a puzzle. Why is entrepreneurship so popular, given the known associated economic penalties and risks? Self-employed business founders earn lower incomes on average than comparable workers in paid employment (Hall & Woodward, 2010; Hamilton, 2000; Hyytinen, Ilmakunnas, & Toivanen, 2013); they also receive lower average risk-adjusted returns on their business capital than investors in public stock markets (Moskowitz, & Vissing-Jørgensen, 2002). As well, failure is commonplace, with one-half of new ventures closing within the first three or four years (Frankish, Roberts, Coad, Spears, & Storey, 2013; Headd, 2003). Incomes are also much more volatile and unpredictable in entrepreneurship than in paid employment (Carrington, McCue & Pierce, 1996; Heaton & Lucas, 2000). Yet high percentages of survey respondents across a variety of countries report that they want to become entrepreneurs – numbers that reach 70% in the USA and 80% in Poland (Blanchflower, Oswald, & Stutzer, 2001).

Several explanations of this puzzle have been proposed, but none completely explain it. These include data measurement problems (Kartashova, 2014; Astebro & Chen, 2014); cognitive biases like over-optimism (de Meza, 2002); love of skewness

(Parker, 2018, chapter 12); and learning benefits (Campanale, 2010). This article explores a hitherto largely overlooked possibility: the positive messaging bias associated with entrepreneurship in the media relative to other career choices.

A society's institutional context is believed to affect the frequency of entrepreneurship (Boettke & Coyne, 2009). Yet while the mainstream media are societal institutions, their possible impact upon entrepreneurship has been little studied. Media act primarily through the cultural-cognitive institutional pillar described by Scott (2014), affecting the shared understandings and commonly held beliefs of economic actors. These institutions establish plausible and even orthodox expectations about careers which shape perceptions and behaviors. Mass media is known to be a powerful vehicle for articulating and defining public perceptions (McCombs & Shaw, 1972; Wanta, Golan, & Lee, 2004) with an impact that is measurable (King, Schneer & White, 2017).

To explore the possible impact of reporting in the mass media upon career choice, we heed the call of George, Haas and Pentland (2014) for greater use of big data in management research. We conducted a large-scale machine-learning-aided analysis of two prominent English language media sources, one North American and the other European – *The New York Times* and the *Financial Times*. These methods are used to identify the emotional polarity ('sentiment') embedded in the texts published by these newspapers, surrounding key words related to new venture careers – 'entrepreneur' and 'founder' – as well as key words related to more established organization careers – 'manager' and 'executive'. A similar analysis was conducted for a set of newer companies led by their entrepreneurial founders and older established companies. Our methodology combines human evaluation of text excerpts, in order to develop an automated sentiment analyzer using machine learning algorithms. This sentiment analyzer was then combined with big data extraction and analysis of that data.

The findings are striking. The relative frequency of positive sentiments surrounding ‘entrepreneur’ and ‘founder’ is much greater than it is for ‘manager’ and ‘executive’. To a slightly lesser degree, the same relationship holds for newer companies led by their entrepreneurial founder(s) when compared to older established companies. We cannot precisely determine the direction of the causation – whether a pre-existing societal sentiment in favor of entrepreneurship underlies the higher relative frequency of positive sentiment in news reporting, or if the news reporting caused the increase in positive sentiment. We suspect and our analysis (partially) supports the belief that, over time, causation flows in both directions. Whether consequence or cause, a sustained period of higher relative frequency of positive sentiment towards entrepreneurship can help explain its popularity, despite the potentially adverse expected economic consequences of this career choice. The mass media appear to play a role in ‘de-risking’ entrepreneurship by inflating the positive feelings associated with it.

METHODS AND DATA

We analyzed the sentiment associated with business-related textual excerpts from sentences extracted from the *New York Times* and *Financial Times* containing certain keywords related to career choices in the business world. ‘Big data’ methods were used to train a nearest-centroid classifier with a dataset of sentences that human participants had assessed. Once trained, this classifier was used to assign a sentiment to over eight hundred thousand excerpts extracted from media articles spanning roughly a dozen years. Positive sentiment rates were defined as the proportion of total excerpts which were classified. Positive sentiment rates were calculated for excerpts containing the career categories ‘entrepreneur/founder’ and ‘executive/manager’. As robustness checks, we conducted a parallel analysis for a set of well-known American newer companies still led by their founders and older established companies led by professional managers using *The New York Times* data.

In what follows, we will first describe the data sources, followed by our methodology based on machine-learning and sentiment analysis.

Data Sources

We adopted a conservative approach towards the selection of our data sources, focusing on two traditional media outlets: *The New York Times* (NYT) and the *Financial Times* (FT). These media outlets preceded and have made the transition into the digital age.[1] They provide a data set with a relatively long historical record, enabling us to track how sentiments may have changed over time. It is important to note that when we get to the era of online news, there is evidence that readers of mainstream media, such as the *NYT* and the *FT*, are quite consistent in their consumption of news, as new formats and user-generated content are not as popular among this group (Boczkowski & Mitchelstein, 2013:4). However, there is no reason, *ex ante*, to believe these traditional media outlets would necessarily be partial towards entrepreneurship, or newer ventures compared to other media. Indeed, if the *NYT* and *FT* have a bias one would expect it to favor established organizations and conventional careers. Although both newspapers have an international readership, the *FT* offers a more British and European view on the world of

[1] New York Times (2014). Innovation. http://www.presscouncil.org.au/uploads/52321/ufiles/The_New_York_Times_Innovation_Report_-_March_2014.pdf.

business, whereas the *NYT* offers an American perspective. Second, it is commonly accepted that both newspapers play an important role as public policy agenda-setters, and accordingly they may affect the broader institutional fabric of their respective societies.

Methodology and Sentiment Analysis

Our methodology comprised three steps: extracting target sentences from both publications; training a classifier to autonomously evaluate our dataset; and performing sentiment analysis on the extracted excerpts.

The NYT API provides access, through HTTP requests, to the articles that match a specific search query. We searched keyword by keyword, and month by month from January 1999 to December 2014, and collected all the articles returned by the API. For each article, we stored relevant fields: publication date, section name, web URL, snippet, abstract, headline and lead paragraph. The four last fields (snippet, abstract, headline, and lead paragraph) were excerpts from the original article that contained the target sentences. We extracted 494,987 text excerpts containing keywords related to jobs in established organizations (‘manager’, ‘executive’), and careers in a new venture (‘entrepreneur’, ‘founder’) from the *NYT* over this period.

For the *FT* we were given access to the full text news dataset from 2003 to 2014 in XML files. These files were provided by the *FT*'s staff and came from the printed version of the newspaper. The first step was to parse these XML files. For each article, we searched every article and collected all the sentences that contained any of the keyword terms. Using the same keywords as for the *NYT*, we retrieved 318,055 sentences. Standard natural language pre-processing – stemming, removal of stop words, conversion to lowercase – were applied to the excerpts from both publications.

To power an artificial intelligence program able to analyze vast amounts of text and assess the related sentiment, we had 190 human subjects report their emotional response to a smaller development dataset of 8,996 sentences collected from the 2013 NYT. These participants completed the sentence assessment task at the Behavioral Lab of a North American business school. Participants went to the Behavioral Lab at a scheduled time and for 60 minutes (with an 8-minute break). They were asked to indicate their emotional reaction to the assessed sentences on a 5-point scale. To achieve redundancy each sentence was evaluated by at least three different participants; as a result, the 8,996 sentences in the development data set were evaluated 48,927 times.[2]

We developed our own automated classifier for several reasons. First, it is important that the phrase/sentence, and not just the word, be the unit of analysis, so that the targeted keywords are analysed in their textual context. The unsuitability of general predefined word lists for the analysis of '10-K text' (i.e. management discussion and analysis) was reported by Loughran and McDonald (2011). They found that word lists developed for other disciplines systematically misclassify common words. In a subsequent review article they noted that "The computational linguistics literature has long emphasized the importance of developing categorization procedures in the context of the problem being studied (e.g., Berelson [1952])." (Loughran & McDonald, 2016: 1208). Indeed, our preliminary analysis of the excerpts using the popular Linguistic Inquiry and Word Count (LIWC: Kahn, Tobin, Massey & Andersen, 2007) did not

[2] The frequency of responses was: +2 = 4405, +1 = 17904, 0 = 11860, -1 = 11597, -2 = 3161 (with -2 being very negative and +2 being very positive).

perform well. When applied to the 8,996 excerpts assessed by human subject with three sentiment categories, positive, neutral and negative, LIWC had an accuracy of just 36%. This was because LIWC classified 52% of excerpts as neutral, while for the human subjects' responses this number was only 11%. When neutral responses were excluded LIWC's accuracy increased to 69%. (See Appendix A for details.) A purpose built classifer was developed to address these concerns.

Sentiment analysis is a classification task, where the classifier (when it is a binary classifier) decides between two classes: positive or negative. We used 'scikit-learn', a Python library for machine learning that provides multiple classifiers and tools for data mining and data analysis. For this task, the feature extractor turned every sentence from the 8,996 sentences in this dataset into a vector that counted the frequency of each word in the sentence. Next, this set of feature vectors was split into two: training and test sets. Since it is difficult to predict, a priori, which classifier will work best (Wolpert, 1996), a variety of classifiers were trained and tested, and the one most accurate for this setting selected.[3] The best classifier was the Nearest Centroid (with Euclidean distance and no shrink threshold) with an accuracy of 0.72 and a Matthews[4] correlation coefficient of 0.40. More details on the performance of this classifier are provided in Appendix B.

RESULTS

Table 1 displays the total number of sentences, grouped by key word category, as well as the number of sentences with positive and negative emotional valence. It also

[3] Some of these classifiers were: Nearest centroid, Ridge, Perceptron, Passive-aggressive, K-Neighbors, Random forest, Linear SVC (support vector machine), SGD (stochastic gradient descent), and Naive Bayes.
[4] https://en.wikipedia.org/wiki/Matthews_correlation_coefficient

gives the corresponding percentages for each category as retrieved from the *NYT* from 1999 to 2014, with corresponding results for the *FT* from 2003 to 2014.

Insert Table 1 about here

When considering the excerpts associated with business and entrepreneurial professions, the results show that the combined terms “manager” and “executive” yield many more appearances in our data than “entrepreneur” and “founder”. Such is the case for both the *NYT* (a proportion of 4.6 to 1) and even more acutely for the *FT* (8.0 to 1). Overall, both newspapers have many more mentions of words related to established business professions than to entrepreneurs and founders. During the time periods studied the coverage of the *NYT* and the *FT* are quantitatively similar when writing about “executives” and “managers” (these professions were mentioned an average of 74.3 and 77.9 times per publication day[5], in each respective newspaper), but their frequencies diverge when mentioning “entrepreneurs” and “founders” (16.1 times/day in the NYT versus 9.7 times/publication day in the FT).

Strikingly, the sentiments associated with occurrences of the key words ‘entrepreneur’ and ‘founder’ were overwhelmingly positive: 81% for the *NYT* and slightly less at 74% for the *FT*. The positive sentiment associated with the key words ‘executive’ and manager” were substantially less, and similar for both publications at 53% for the *NYT* and 54% for the *FT*. Moreover, these differences are statistically significant. Statistical tests of difference need to take account of the precision of the different scores. To do so we constructed ‘confusion matrices’ for both career categories

[5] The *NYT* publishes 7 days a week; the *FT* Monday through Saturday.

for the *NYT* and the *FT*. Sentiment differences between career categories for both the *NYT* and the *FT* are highly significant. See Appendix C for details.

Insert Figure 1 about here

As shown in Figure 1, monthly frequencies have fluctuated over time for both categories. The *NYT* witnessed a sharp increase in reports including the combined frequency of ‘manager’ and ‘executive’ starting in 2006 and continuing well into 2012. As depicted in Figure 1, the frequency of ‘entrepreneur’ and ‘founder’ for the *NYT* increased since 2006, with a spike in 2012. For the *FT* these terms exhibited a modest increase over the reporting period. Figure 1 also includes 95% confidence intervals for the frequency time series derived using a 6-month rolling window.

Insert Figure 2 about here

Figure 2 shows the percentages of positive responses for the keyword categories for the *NYT* and the *FT* during the reporting periods. Even though there are month-to-month variations, the overall sentiment associated with both keyword categories is consistent for both publications over the study period. Noting the use of 95% confidence intervals, the rates of positive sentiments for entrepreneurs/founders exceed those of managers/executives both for the *NYT* and the *FT* throughout the sample period.

We next conduct some follow-on empirical analysis to probe the results in an abductive manner, to explore the validity of possible underlying mechanisms.

‘Cause and Effect’

Are positive sentiments statistically associated with subsequent participation in entrepreneurship – and/or vice-versa? To explore this question, we estimate a vector autoregression (VAR) model in which a) the rate of positive sentiments is related to lagged rates of entrepreneurship; and b) the rate of entrepreneurship is related to lagged rates of positive sentiments. The coefficients of a VAR identify systematic patterns of interdependence between multiple time series, while taking account of 'state dependence' whereby past values of dependent variables also correlate with future values of those same dependent variables (Henley, 2004; Andersson & Koster, 2011).

The time series of interest for our research were operationalized as: the ratio of positive to total entrepreneurship sentiments time series (PS) derived from the *NYT* and *FT*; and the self-employment rate time series (SE) for both the US and the UK respectively. Appendix D provides more details about the data and analysis. Table 2 reports the results of 'Granger causality' tests derived from the VAR, which test the joint significant of lagged variables on current ones. The results show that for the US (but not the UK) the relationship between past positive media sentiments and future rates of entrepreneurship appears to be statistically significant. For both countries, Granger causality also flows the other way, from past rates of entrepreneurship to future positive sentiments. While suggestive, we stress that not too much should be read into these findings. Granger causality is simply a form of predictive causality, which neither identifies underlying mechanisms nor establishes behavioral causality. These results must therefore be treated with caution, being unable to rule out behavioral causality running in one direction, the other direction, or in both directions at the same time.

Insert Table 2 about here

Impact of Specific Exogenous Events

Since media reporting responds to real world events, we next explore whether changes in sentiments are associated with specific identifiable events. We sample a limited number of different types of events including economic, managerial and innovative. We would characterize the Enron scandal as an example of a US managerial event; the onset of the Great Recession is an example of an economic event; and the launch of the iPhone as an innovative event. To this end, we adopted a two-step empirical approach. First, a research assistant surveyed the two newspapers over the sample periods to identify consequential events, together with their dates. Second, taking each event as an exogenous shock, we compared positive sentiments towards entrepreneurs relative to managers before and after their occurrence. Specifically, we conducted a 'difference-in-difference-in-difference' (DDD) analysis, comparing post- and pre-event differences in net positive sentiment between entrepreneurs and managers. Details of this analysis are reported in Appendix E. Tables 3 and 4 present the results for the US and UK.

Insert Table 3 about here

Table 3 presents the results for nine events in the US within the sample period. Two related events stand out: the onset of the Great Recession in September 2008 and relatedly the automakers' bailout two months later. Both of these events are associated with significant shifts in the ratio of positive sentiment of entrepreneurs relative to managers, most strongly in the case of the Great Recession. The more challenging economic conditions associated with the Great Recession and the bailouts of some big banks, insurance and auto companies may have resulted in disillusionment with large established firms, as well as possibly increasing the positive sentiment towards entrepreneurs and founders – thus reducing the relative positive sentiments towards managers and executives. Also notable for being insignificant was the bursting of the dot com bubble in April 2001. Even though the dot.com bubble was associated with many innovative entrepreneurial ventures, the bursting of this bubble does not appear to have negatively affected sentiment towards entrepreneurship. The other more political event that seemed to have had a (marginally) statistically significant impact on relative sentiments was the 'Occupy Wall Street' protests. These demonstrations drew attention to inequality and the undue influence on government of corporations, especially those in the financial services sector. Interestingly, none of the single enterprise events were associated with significant changes in overall sentiments. For example, the Enron scandal which broke in October 2001 had no significant association with changes in manager/executive sentiment. Likewise, neither of Apple's product innovation announcements were significantly associated with changes in entrepreneur/founder sentiment.

Insert Table 4 about here

Table 4 presents results for the UK. Given the importance of the US to the global economy, several events which were part of the US study also appear in the UK list. However, of these US-centric events only one – the launch of the iPhone – is associated with a significant increase in positive sentiments of entrepreneurship relative to managers. Several UK and European events did have an impact on sentiments associated with FT reporting. The demise of the last British domestic volume automaker, MG Rover, and the first bank run in Britain for over a century (which led to the collapse of a large bank) were both associated with increases in positive relative sentiments towards entrepreneurship. Similar effects were observed for a critical moment in the 2011 Euro crisis (the collapse of sovereign Greek credit) and the LIBOR interest rate rigging scandal. Interestingly, the Great Recession did not have a similar effect in the UK as it did in the US. It may be that the reporting had already reacted to the preceding financial crisis beginning in early 2008.

In summary, these results suggest that sentiments reported towards entrepreneurship and managers may respond to some specific events. Notably the *FT* appears to be more sensitive to business related events than the *NYT*. For the *FT* five of seven events had a significant association with change in relative sentiment; for the *NYT* it was two of nine events. This result may be an artifact of the somewhat different orientations of these publications. The *NYT* is a general news publication with a business component; the *FT* is a business/finance publication with a general news component. It is important to keep in mind that the sentiments being reported are derived from the journalists writing for the *NYT* and *FT*. It is plausible that the reporting by the *FT* pool of journalists would be more sensitive to business events than the journalists of the *NYT*.

Sentiments Associated with Newer Founder-led Companies

We next explore whether the difference in emotional valence associated with the different career descriptors also hold for newer companies led by their founders compared to established companies led by professional managers. Here we report the results for the FANG companies (Facebook, Amazon, Netflix and Google, now Alphabet). Amazon, Netflix and Google existed during the entire period of this study; Facebook came into existence in 2004. Of this group, Amazon was the only publicly traded company at the beginning of the study period; all were public by the end of this period. The founder(s) still played significant roles in these companies during the study period. Since these companies are all based in the United States, we only report results for the *NYT*. While we used the entire study period (1999 to 2014) the occurrences of all four FANG companies did not consistently surpass 500 per month until 2009. In the early years, occurrences were dominated by Amazon.

To establish a comparator group of established companies we took the largest companies from Fortune 500 list for the start of our study period, 1999. The companies, in order of sales revenue size, were General Motors, Wal-Mart Stores, Exxon Mobil, Ford Motor and General Electric. In order to be conservative with this comparison we selected only companies that were 'admired'. All these large companies, except for GM, were also Fortune's 'Most Admired' company in their industry sub-group for the start of the study period (1999). For this reason, and to have only one auto company in this comparison, Ford Motor was included, and GM excluded, from the comparator group.

Insert Table 5 about here

Table 5 reports the aggregate results for each company. While not as pronounced, the results from the company categories are directionally similar to, and supportive of, the results for the key word categories. That is, newer ventures led by their entrepreneurial founders had a higher proportion of positive sentiment than did older established companies led by professional executives. Interestingly the four FANG companies had more occurrences over the entire study period than did the four large Fortune companies during the same period, even though many of the FANG companies did not exist during the earlier years. Exxon Mobil reported the lowest positive sentiment results for any of the reported companies (36%). This is suggestive of a possible industry effect upon sentiment (an interesting topic for future study).

Figure 3 presents the aggregate percentage positive sentiment over the study period, using a 6-month moving average to smooth month-to-month variability. The difference between the two categories appears to be consistent over the study period, (except perhaps 1999-2000, when the FANG group was represented only by Amazon). Overall, we conclude that our earlier results about higher positive sentiments among entrepreneurs are robust to different notions of entrepreneurship based on innovation and scale, though we acknowledge that the FANG companies are less ‘entrepreneurial’ in other respects, e.g. by no longer being either ‘young’ or ‘small’.

Insert Figure 3 & Table 6 about here

By the end of our study time period all the FANG companies were publicly traded and had experienced large increases in their share values. This raises the possibility that the higher positive sentiments for these companies simply reflects their growing stock prices. To explore this possibility, we calculated the correlations between their PS and changes in their stock prices (see Appendix F for details). Table 6 presents the results. Most of the correlations are modest and generally close to zero. An exception is

Facebook, with two large and significant pairwise correlation coefficients. However, after adjusting for conducting multiple hypothesis tests, using a Bonferroni correction, even the correlation coefficients for Facebook become marginally insignificant. None of the other correlation coefficients in Table 6 are significant using the Bonferroni correction, leading us to rule out rising stock prices as a simple explanation of the superior positive sentiments enjoyed by the FANG companies relative to established firms.

DISCUSSION

Mainstream media are important societal institutions and are believed to influence their economic and social environments. Employing a big data approach, this research measured the emotional valence or sentiment associated with excerpts from *The New York Times* and the *Financial Times* that mentioned two different broad occupational categories. Using longitudinal micro data, this research uncovered a relative positive sentiment reporting bias in favor of entrepreneurs and founders relative to executives and managers, as well as of newer founder-led companies relative to established companies.

More research needs to be done to establish exactly how media sentiment affects career choice. However, we believe the results of this research point to an explaination for the popularity of becoming an entrepreneur, despite the lower expected economic rewards associated with these careers. Dan Lovallo and Daniel Kahneman referred to this in terms of “delusions of success” (2003). Lovallo and Kahneman (2003) assert that both managers and entrepreneurs are overly optimistic and go on to identify several possible sources for this optimism – exaggeration of their own talents, the illusion of being able to control the uncontrollable, and anchoring and organizational pressures. This “entrepreneurial delusion” may be, at least in part, created by ‘biased’ reporting in the mainstream (and other) media. Nightingale and Coad (2014) speculated that the popular press accords disproportionate visibility to a few entrepreneurial winners, which may lead to excessive entrepreneurial entry.

Extending this reasoning, we assert that the perception of riskiness is not just related to the frequency of mentions or familiarity – the so-called “availability heuristic” (Kahneman, 2012). The associated sentiments are also meaningful. Slovic, Finucane, Peters and MacGregor (2002) proposed an affect heuristic, whereby peoples’ likes and dislikes determine their beliefs about the world. They assert that people commonly form opinions and make choices that directly express their feelings and their basic tendency to approach or avoid, often without being aware they are doing so. Affect, emotion and sentiment are related constructs. In a compelling demonstration of the affect heuristic, Slovic’s research team surveyed opinions about various technologies, including water fluoridation, chemical plants, food preservatives, and cars, and asked their respondents to list both the benefits and the risks of each technology. They observed an implausibly high negative correlation between estimates of the level of benefit and the level of risk respondents attributed to the technologies. People positively disposed toward a technology rated it as offering large benefits and imposing little risk; when they disliked a technology, they could think only of its disadvantages, and few advantages came to mind. Remarkably, even members of the British Toxicology Society responded similarly.

A similar effect may be at work with careers. A known risk – the (low) probability of a new business venture becoming economically successful – is assessed by prospective new business founders as less salient because they associate more positive sentiments to this career than they do to a (less economically risky) managerial career in an established company. If our findings are generally representative of the media, and we suspect they are, then these inclinations may be induced and reinforced by the positive sentiment biases toward entrepreneurship and new ventures present in media outlets.

A consequence of this bias may be additional entrepreneurial entrants: possibly more than can be profitably sustained. While the outcome would be economically deleterious for the average entrepreneurial entrant, the effect on the overall economy could still be positive. Dosi and Lovallo (1995) call failed entrepreneurs that signal new markets to more qualified competitors "optimistic martyrs" – good for others and the overall economy but bad for themselves (and their investors). Entrepreneurial activity is believed to confer other positive externalities that increase overall welfare (Becker and Murphy, 2000; Bernardo & Welch, 2001; Nordhaus, 2004). Thus, the *NYT* and the *FT*, along with other media, may be promulgating a 'meme' (Dawkins, 1989; Blackmore, 1999) – a positive sentiment bias favoring an entrepreneurial career choice – that while economically detrimental to the average business founder or entrepreneur, improves overall economic welfare within their societies. Following the logic of group selection, the mass media could help to increase the frequency of this meme (Boyd & Richerson, 2002), and, up to a point, societies with a higher frequency of this meme would outperform and economically outcompete societies with a lower frequency of this meme.

On the other hand, some research suggests that certain societies may have too much entrepreneurial entry, leading to wasted resources (de Meza, 2002; Henrekson & Sanandaji, 2014). If this is the case a positive sentiment meme would exacerbate the over-investment associated with excessive entrepreneurial entry, and the mass media by propagating a positive entrepreneurial meme would be acting against the interests of that society. It is plausible that an ‘optimal’ level of entrepreneurial activity exists in any society; yet we lack a good sense of what this ‘optimal’ level may be in practice. Until this gap in our knowledge is filled, we can only speculate about the impact of the media promoting more positive sentiments towards entrepreneurship.

The implications for public policy are nuanced (Shane, 2009). Even if the high rate of entrepreneurial entry induced by pro-entrepreneurial media bias makes the average entrepreneur/founder economically worse off, but society as a whole better off, should policy-makers encourage citizens to engage in a privately risky activity albeit with positive benefits to the overall economy (Nightingale & Coad, 2014)? There is no clear answer since one’s position on this question depends on their view about social welfare and the proper role of government. For example, a utilitarian might well countenance a pro-entrepreneurship policy that knowingly imposes costs on some citizens while benefiting other citizens more – at the same time a classical liberal would oppose this position. In practice, a common policy stance in many countries is to use public money to promote entrepreneurship, through a variety of initiatives and interventions (Parker, 2018, Chapters 18-21). One can question whether this is a wise use of public resources, even if entrepreneurship generates positive externalities, since the downsides of business failure may hit some vulnerable populations disproportionately (de Meza, 2002).

This article also carries potential implications and perhaps dilemmas for business educators, policymakers and the media. Regarding business education, there is a widespread belief among many educators that entrepreneurship is the key to economic growth and prosperity – and entrepreneurship courses often promote this career choice (Torrance et al., 2013). There is a danger that educators who encourage entrepreneurship by lauding examples of highly successful start-ups while ignoring or diminishing the risk of new venture failure contribute to their students forming a misleading sense of the difficulties and risks associated with this career choice. And media bias reinforces this perception. As a result, many students could eschew careers in paid employment in pursuit of their uncertain entrepreneurial aspirations. We believe educators have an obligation to take a balanced and realistic approach, drawing students' attention to known risks associated with start-ups as well as possible pro-entrepreneurship media biases, when they are weighing their future career possibilities.

To conclude, we have discovered a positive sentiment bias towards entrepreneurship within two mainstream media outlets in two different countries. Thus while the news these media report is not fake, it is biased. Biased because it reports a sentiment towards entrepreneurship that does not reflect the economic reality related to an entrepreneurial career choice. We do not know how widespread such positive sentiment biases towards entrepreneurship may be. The methodology employed in this research can be extended to other media, as well as other countries, cultures and languages. Doing so would allow the examination of the differences in sentiment biases across different types of media as well as different cultural and language groups.

APPENDIX A

Linguistic Inquiry and Word Count (LIWC) is the de facto standard for sentiment analysis. LIWC contains a dictionary of 4,500 words classified in different predefined categories. Many of the LIWC categories are arranged hierarchically. For instance, all anger words, by definition, are categorized as negative emotion and overall emotion words.

We tested this tool on the same 8,996-sentence datasets assessed by our participants. For this test, we describe our participants' assessment process and LIWC's scores, and then compare the results obtained by the two processes.

Each of the 190 participants scored a random subset of the 8,996 test sentences as very negative (-2), negative (-1), neutral (0), positive (+1), very positive (+2). Not every participant assessed all the sentences, but every sentence was evaluated by at least three different participants (i.e. we set a redundancy greater or equal to three). The final score for each sentence was the average of its individual scores. Therefore, every sentence got a final value between -2 and +2. Finally, this value was normalized to the range [-1, +1].

LIWC analyzes every sentence word by word. As each target word is processed, the dictionary is searched, looking for a dictionary match with the current target word. If the target word matches the category word, the appropriate word category scale/s for that word is/are incremented. Due to its hierarchical category arrangement, only the two top categories for emotions, posemo (positive emotion) and negemo (negative emotion), were considered for this comparison (the other categories for emotions are already contained in the two top categories). Upon analysis completion, every sentence got a value for posemo, between 0 and 100 (%), and another value for negemo, also between 0 and 100 (%). This two metrics were used to calculate the final score for each sentence as follows:

if negemo < posemo, then score = posemo

if negemo > posemo, then score = -negemo

if negemo = posemo, then score = 0

Finally, this value was normalized to the range [-1, +1]. (The described formula was chosen to favour LIWC; e.g. the alternative formula score = posemo – negemo was discarded because it gave worse results for LIWC.)

Comparison

Comparing LIWC results with participant scores revealed some important differences. Figure A1 illustrates the differences in the distributions. LIWC is strongly biased towards 0 (it evaluates numerous sentences as neutral). Relatedly LIWC classifies fewer sentences as more negative or more positive, in comparison to the participants' scores. A "two proportions z test" between these distributions yielded a test statistic of z = 10.93 and a p-value of 0.00. This test was run after removing the sentences with a score equal to 0, which makes the two distributions more similar, but still significantly different as the test shows. (Again, this decision was also made to favour LIWC: retaining the 0s led to even larger differences between the two distributions.) For the 8,996 sentences assessed by human subjects, LIWC had the same polarity as the participants in 3,202 instances (and different polarity in 5,794). Overall, LIWC's accuracy was only 36%. However when non-neutral excerpts were excluded LIWC did better.

LIWC classified 4640 excerpts (52%) as neutral; 4111 contained no emotion words, neither positive nor negative; for the remaining 549 excerpts the number of positive emotion words and negative emotion words were equal. Human subjects assessed 1020 excerpts as neutral (11%). Combined there were 5127 neutral excepts. For the 3,869 non-neutral excepts, 1,053 were assessed as negative by human subjects; LIWC agreed in 658 instances. Similarly, humans assessed 2,816 as positive and LIWC agreed in 2,011 instances. For this subset of non-neutral excerpts LIWC had an accuracy of ((658+2011)/3869=) 69%.

APPENDIX B

Accuracy is a global metric for both positive and negative categories. To answer the question ‘What proportion of sentences did the classifier classify correctly?’, let S be the total number of sentences, P the number of real positive sentences as assessed by the participants, and N the number of real negative sentences.

We have:

S = P + N

Let TP (true positive) be the number of real positive sentences that were correctly classified as positive by the classifier, FP (false positive) the number of real negative sentences that were incorrectly classified as positive, TN (true negative) the number of real negative sentences that were correctly classified as negative, and FN (false negative) the number of real positive sentences that were incorrectly classified as negative. Then

P = TP + FN

N = TN + FP

Accuracy is defined as the number of correct predictions divided by the total number of predictions:

A = (TP + TN) / (TP + TN + FP + FN)

The classifier classified correctly 72% of the sentences.

Precision is different for each category. For the positive category (the definition is analogous for the negative category), it answers the question 'What proportion of sentences classified as positive are real positive sentences?' Precision is defined as the number of real positive sentences that were correctly classified as positive divided by the number of total sentences classified as positive (both correctly and incorrectly):

$$P = TP / (TP + FP)$$

The classifier has a precision of 79% when classifying the positive category and 61% for the negative category.

Recall is also different for each category. For the positive category (the definition is analogous for the negative category), it answers the question 'What proportion of real positive sentences are classified as positive?' Recall is defined as the number of real positive sentences that were correctly classified as positive divided by the number of real positive sentences:

$$R = TP / (TP + FN)$$

The classifier has a recall of 76% when classifying the positive category and 64% for the negative category.

F1-score is a measure that combines precision and recall. It is defined as:

$$F1 = 2 \times (P \times R) / (P + R)$$

Accuracy is equal to the average/total recall. Accuracy is usually the starting point to evaluate a classifier performance, but alone typically provides insufficient information. Precision and recall complement accuracy when evaluating a classifier.

Table B1 displays scores for accuracy, average/total precision, recall, and F1. Even though they have the same total value, these are different metrics.

Insert Table B1 about here

For this study, 72% for precision and recall can be regarded as a good result. Some machine learning studies do report higher performance (over 90% in some cases) – but for problems such as facial recognition, where no outside effects or confounding variables can influence the pixel-based classification of the picture. All the information needed for the classification task is available in the picture, so a very high classification accuracy can often be reached in these problems. Sentiment analysis is often applied to movie or book reviews that contain explicit words such as “good”, “bad”, “great”, “terrible”. Other studies are based on online human actions in social media like “thumbs up” and “thumbs down”, or on the presence of emoticons like “:-)” and “:-(”. All these tokens provide the classifier valuable hints for learning to predict when a review or an opinion is positive or negative. These types of sentiment analysis also often report higher accuracy (over 80%).

The experiment reported in this study makes accurate classification a more demanding task. The NYT and FT pride themselves on doing objective journalism. They tend to avoid using words like good or bad and exclude emoticons. Thus, our dataset lacks some of the strong hints available to other classifiers. Our sentences simply contain a business-related word, and when assessing a sentiment polarity to a sentence, our participants only have that keyword and its context in the sentence. We do not have all possible information at the participants’ disposal, only some data that represent the assessor's behavior at that time. We can capture neither every influencing factor nor the assessor's personal circumstances. For these reasons, we regard an accuracy of 72% (supported by an equal precision and recall) as a satisfactory classification result for our problem.

In fact, this value falls into the range deemed acceptable for any classifier (60% and above). As an illustration, consider the following examples of studies on sentiment classification with similar scope and accuracy as ours. First, Chiong et al. (2018) proposed a sentiment analysis-based approach for financial market prediction using news disclosures. Their prediction model scored a maximum accuracy of 59%. Second, Dridi, Atzeni, and Reforgiato-Recupero (2018) applied a classifier to financial news in order to identify bullish and bearish sentiments associated with companies and stocks. They reported an accuracy of 72%. Third, Wilson, Wiebe, and Hoffmann (2005) presented an approach to phrase-level sentiment analysis with an accuracy of 76%.

Fourth, Barnes, Klinger and Schulte im Walde (2017) applied state-of-the-art models for sentiment analysis to several datasets from different domains: Stanford Sentiment Treebank (movie reviews), OpeNER (hotel reviews), SenTube (two datasets of texts taken from YouTube comments regarding automobiles and tablets), and SemEval (tweets collected for the 2013 Semantic Evaluation Workshop). Some of these state-of-the-art models reached an accuracy of over 80% for the movie and hotel review datasets. The two most similar domains to ours were automobile and tablets. For these two datasets, none of the models surpassed 70% accuracy (although these datasets included the neutral class, which we discarded from our dataset, and this usually leads to a lower accuracy than in binary classification).

APPENDIX C

Based on the predicted values and the different precision scores for each class (79% and 61% for the positive and negative class, respectively: see Table B1), we constructed ‘confusion matrices’ of both occupations for both the NYT and the FT and estimated the true values of the negative and the positive classes. See Table C1.

Insert Table C1 about here

For the NYT, the total excerpts containing the keywords “executive” and “manager” was 406,757 (n_1); the number of positive excerpts was 245,106 (y_1); hence

$$p_1 = y_1 / n_1 = 0.60$$

The total excerpts containing the keywords “entrepreneur” and “founder” was 88,253 (n_2); the number of positive excerpts was 62,972 (y_2); hence

$$p_2 = y_2 / n_2 = 0.71$$

$$p = (y_1 + y_2) / (n_1 + n_2)$$

H_0: $\mu_1 = \mu_2$ (the proportions are the same; their means are the same)

H_1: $\mu_1 \neq \mu_2$ (the proportions are different; their means are different)

The two proportions z test provided by the scientific library Statsmodel for Python gave a test statistic of $z = 51.632 > 1.96$, rejecting the null hypothesis in favor of the alternative hypothesis. Thus, the mean differences between the career categories for the NYT are significantly different. The same analysis for the FT data also produced statistically significant results ($z = 28.915 > 1.96$).

APPENDIX D

After estimating the VAR model, we explored whether one variable 'causes' another in the sense of Granger (1969) – i.e. whether the lags of one variable are jointly significant in terms of their association with future values of the other. These 'Granger causality tests' use χ^2 statistics whose null hypothesis is independence of two or more time series. High values (i.e. low *p* values) lead one to reject the null and infer Granger causality from the lagged values of one variable to subsequent values of the other.

To implement the VAR and Granger causality tests, we adopted standard practice and operationalized 'entrepreneurship rates' as national self-employment rates. The US self-employment rate was defined as the ratio of the total number of non-agricultural self-employed unincorporated workers to total non-farm payrolls. These data were not seasonally adjusted, in order to be consistent with the sentiments data. The monthly data series, for January 1998 through to December 2015, was downloaded from the Federal Reserve Bank of St Louis website.[6] The UK self-employment rate was defined as the ratio of the total number of self-employed workers to the total workforce (not seasonally adjusted). The monthly data series for January 2002 through to December 2015, was downloaded as Table EMP01 from the Office of National Statistics website.[7]

For both countries, PS denotes the ratio of positive to total sentiments, and SE denotes the self-employment rate. As is conventional practice, the lag length of the VAR in each country was chosen to minimize the Akaike Information Criterion (AIC). This

---

[6] https://fred.stlouisfed.org

[7] https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/datasets/fulltimeparttimeandtemporaryworkersnotseasonallyadjustedemp01nsa

gave rise to an optimal lag of 6 months in the US case, and 9 months in the UK case. Table D1 presents the estimates of the VAR models while Table 2 (in the text) presents the Granger causality statistics.

Insert Table D1 about here

For each country, Table D1 is divided into two parts, the first reporting results for PS as the dependent variable, and the second reporting results for SE as the dependent variable. For the United States, Panel A shows that there is some limited state dependence for positive sentiments, but no individually significant association between lagged self-employment rates and current positive sentiments. However, collinearity between the lagged independent variables might be obscuring joint significance (tests of which appear in Table 2). Panel B displays two significant positive and one significant negative coefficient for lagged positive sentiments on current US self-employment rates. The UK case displays one significant positive and one significant negative coefficient for lagged self-employment rates on current UK positive sentiment rates. However, there are no significant effects of lagged positive sentiments on UK self-employment rates.

APPENDIX E

Suppose an event happens at time *t*. Let e^{+}_{t+1+j} denote the proportion of positive sentiments relating to entrepreneurship at periods j + 1 = 2,3,… after the event; and let e^{-}_{t+1+j} be the corresponding proportion of negative sentiments relating to entrepreneurship. Define m^{+}_{t+1+j} and m^{-}_{t+1+j} likewise for positive and negative sentiments towards managers. The post-event double difference $D = \left(e^{+}_{t+1+j} - e^{-}_{t+1+j}\right) - \left(m^{+}_{t+1+j} - m^{-}_{t+1+j}\right)$, which is the difference between net positive sentiments towards entrepreneurs and managers, is then compared with the corresponding pre-event double difference to form the DDD estimator:

$$DDD = \left[\left(e^{+}_{t+1+j} - e^{-}_{t+1+j}\right) - \left(m^{+}_{t+1+j} - m^{-}_{t+1+j}\right)\right] - \left[\left(e^{+}_{t-1-j} - e^{-}_{t-1-j}\right) - \left(m^{+}_{t-1-j} - m^{-}_{t-1-j}\right)\right]$$

Standard errors for *DDD* can be estimated, and a *t* test of the mean difference performed. A range of *j* values capturing up to 24 months each side of the event (dropping the most proximate month either side to reduce the risk of overlapping temporal events) were used in the estimations.

APPENDIX F

Consider the following pairwise correlation coefficients for all companies *j*:

$$c_j = corr(PS_{jt}, \Delta \ln s_{jt})$$

where PS_{jt} is the positive sentiment rate for company *j* at time *t*, and $\Delta \ln s_{jt}$ is the change in log stock prices for company *j* at *t*, (i.e. the growth rate in the stock price). A version of c_j where ΔPS_{jt} is used instead of PS_{jt}, was also computed: this coefficient is denoted by c_j^{Δ}. If the positive sentiment level covaries strongly with stock price growth rates, large positive values of c_j should be observed, which would be clustered closer to 1 than to 0. Similar results would be expected if *changes in sentiment* were correlated with changes in share price. Contemporaneous correlations were computed because the efficient market hypothesis suggests that all information is quickly impounded in prices. We also computed this correlation with a one-period lagged impact on the share price $c_j = corr(PS_{jt}, \Delta \ln s_{jt+1})$ – see the final column of Table 6 (in the text).

Data on monthly stock prices for each company were downloaded from Yahoo Finance. The sample period for a given company was the longest available continuous time span, post-IPO, within the December 1999 – December 2014 window. The maximum window was usable for Amazon (whose IPO was in 1997) and the established companies; but the sample periods were shorter for Netflix, Facebook and Google, which had IPOs in May 2002, May 2012 and August 2004, respectively.

TABLE 1

General Sentiment Polarity by Key Word Category

Media Outlet	Key Words	total occurrences	negative #	negative %	positive #	positive %
The New York Times	executive/manager	406,757	190,578	47%	216,176	53%
	entrepreneur/founder	88,253	16,869	19%	71,384	81%
Financial Times	executive/manager	282,724	130,238	46%	152,486	54%
	entrepreneur/founder	35,331	9,262	26%	26,069	74%

TABLE 2

Granger causality tests

	United States	United Kingdom
SE $\rightarrow$ PS	$\chi^2(6) = 15.59$**	$\chi^2(9) = 19.52$**
PS $\rightarrow$ SE	$\chi^2(6) = 18.75$***	$\chi^2(9) = 6.18$

***: Significant at 0.01; **: Significant at 0.05

TABLE 3

Difference-in-Difference-in-Difference results for the United States
24 months each side of the event ± 1 month each side

Event	Date	Mean DD pre	Mean DD post	t statistic	p value
Dot com bubble bursts	4/2001	0.529	0.522	0.42	0.679
Enron scandal	10/2001	0.521	0.539	1.28	0.208
Steve Jobs launches iTunes	4/2003	0.521	0.547	1.64	0.109
Steve Jobs launches iPhone	1/2007	0.560	0.559	0.08	0.939
Great Recession	9/2008	0.544	0.576	2.94	0.005***
Automakers bailout	11/2008	0.547	0.570	2.02	0.049**
Deepwater Horizon oil spill began	4/2010	0.574	0.564	0.67	0.506
"Occupy Wall Street" protests began	9/2011	0.555	0.585	1.74	0.089*
Tesla delivers first Model S sedan	6/2012	0.563	0.555	0.41	0.683

Degrees of freedom are 46 in all cases.
***: Significant at 0.01; **: Significant at 0.05

TABLE 4

Difference-in-Difference-in-Difference results for the United Kingdom
24 months each side of the event ± 1 month each side

Event	Date	Mean DD pre	Mean DD post	t statistic	p value
Richard Branson signs deal for space tourism company †	9/2004	0.353	0.358	0.43	0.670
MG Rover (auto maker) goes into administration	4/2005	0.349	0.367	1.96	0.056*
Steve Jobs launches iPhone	1/2007	0.361	0.411	4.08	0.000***
Bank run leads to collapse of Northern Rock	2/2008	0.374	0.433	5.54	0.000***
Great Recession	9/2008	0.389	0.402	0.87	0.389
Stock markets fall sharply on prospect of a Greek default	6/2011	0.373	0.423	3.12	0.003***
LIBOR rate setting scandal exposed ‡	7/2012	0.391	0.425	1.86	0.070*

Degrees of freedom are 46 in all cases except † (df = 40) and ‡ (df = 44)

***: Significant at 0.01; **: Significant at 0.05

TABLE 5
General Sentiment Polarity by Company and Category
1999 to 2014

Company	Total occurrences	Negative #	Negative %	Positive #	Positive %
Facebook	33627	9497	28%	24130	72%
Amazon	17335	4914	28%	12421	72%
Netflix	3791	866	23%	2925	77%
Google	33932	8585	25%	25347	75%
Category total	88685	23862	27%	64823	73%
Wal-Mart	12901	5169	40%	7732	60%
Exxon Mobil	4382	2793	64%	1589	36%
Ford Motor	5423	1632	30%	3791	70%
General Electric	7525	2658	35%	4867	65%
Category Total	61640	22073	36%	39567	64%

TABLE 6
Relationship between Positive Sentiment and Changes in Share Price

Company	c_j	c_j^{Δ}	$corr(PS_{jt}, \Delta \ln s_{jt+1})$
Facebook	0.187	0.505***	-0.492***
Amazon	-0.001	0.034	0.024
Netflix	0.178**	0.151*	-0.177**
Google	0.103	0.096	0.058
Walmart	-0.042	-0.025	0.108
Exxon Mobil	0.167**	0.093	0.038
Ford	0.096	0.054	-0.013
General Electric	0.104	0.070	-0.036

Single tests: ***: Significant at 0.01; **: Significant at 0.05; *: Significant at 0.10.

Multiple tests: Bonferroni corrections divide these values by 8x3=24, e.g. 0.01÷24 = 0.0004. None of the coefficients in the table achieve significance at this level.

TABLE B1
Classification Report

	Precision	Recall	F1-score
Negative	0.61	0.64	0.63
Positive	0.79	0.76	0.77
Average/Total	0.72	0.72	0.72

TABLE C1
Confusion Matrices for Careers by Publication

Publication	Career		True -	True +	Total Predicted
The New York Times	manager/executive	Predicted -	116,254	74,327	190,581 (47%)
		Predicted +	45,397	170,779	216,176 (53%)
		Total True	161,651 (40%)	245,106 (60%)	406,757 (100%)
	entrepreneur/founder	Predicted -	10,290	6,579	16,879 (19%)
		Predicted +	14,991	56,393	71,384 (81%)
		Total True	25,281 (29%)	62,972 (71%)	88,253 (100%)
Financial Times	manager/executive	Predicted -	79,445	50,793	130,238 (46%)
		Predicted +	32,022	120,464	152,486 (54%)
		Total True	111,467 (39%)	171,257 (61%)	282,724 (100%)
	entrepreneur/founder	Predicted -	5,650	3,612	9,262 (26%)
		Predicted +	5,474	20,595	26,019 (74%)
		Total True	11,124 (31%)	24,207 (69%)	35,331 (100%)

TABLE D1
VAR Estimates

	United States		United Kingdom	
	A. PS equation		*A. PS equation*	
	β	ese(β)	β	ese(β)
PS(-1)	0.186***	0.072	1.138***	0.088
PS(-2)	0.116	0.073	-0.174	0.134
PS(-3)	0.102	0.073	-0.660***	0.129
PS(-4)	-0.023	0.073	0.527***	0.134
PS(-5)	-0.043	0.073	0.034	0.143
PS(-6)	-0.018	0.073	-0.372***	0.133
PS(-7)			0.384***	0.125
PS(-8)			-0.004	0.129
PS(-9)			-0.125	0.085
SE(-1)	0.535	1.445	2.900**	1.143
SE(-2)	-1.357	1.875	-1.891	1.643
SE(-3)	-0.172	1.736	-3.225**	1.635
SE(-4)	1.924	1.736	2.146	1.681
SE(-5)	-0.114	1.843	-0.051	1.779
SE(-6)	-3.064	1.439	1.814	1.752
SE(-7)			-1.182	1.751
SE(-8)			-2.353	1.772
SE(-9)			1.857	1.219
Constant	0.702***	0.133	0.184 ***	0.069
R^2	0.212		0.799	
χ^2	50.14***		509.55 ***	

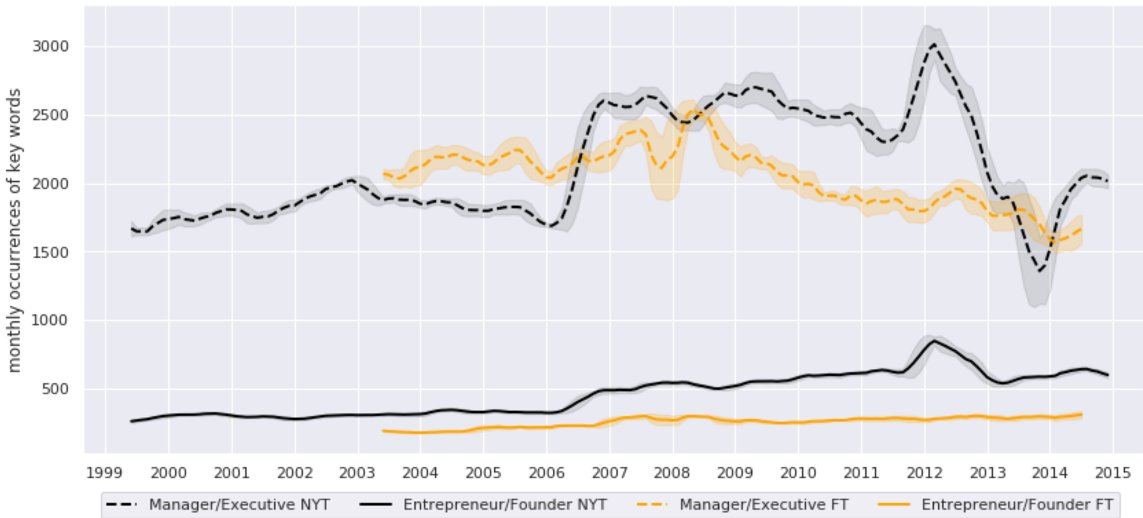
	B. SE equation		*B. SE equation*	
	β	ese(β)	β	ese(β)
PS(-1)	-0.008**	0.004	0.010	0.007
PS(-2)	-0.001	0.004	-0.005	0.011
PS(-3)	-0.001	0.004	-0.003	0.010
PS(-4)	-0.006*	0.004	0.004	0.011
PS(-5)	0.008**	0.004	-0.006	0.011
PS(-6)	-0.009**	0.004	0.011	0.010
PS(-7)			-0.005	0.010
PS(-8)			-0.006	0.010
PS(-9)			0.008	0.007
SE(-1)	0.887***	0.073	1.031***	0.090
SE(-2)	0.050	0.094	-0.036	0.130
SE(-3)	0.079	0.087	-0.200	0.129

SE(-4)	-0.406***	0.087	0.437***	0.133
SE(-5)	0.230**	0.092	-0.246*	0.140
SE(-6)	0.085	0.072	-0.156	0.138
SE(-7)			0.172	0.138
SE(-8)			-0.193	0.140
SE(-9)			0.209**	0.096
Constant	0.018***	0.007	-0.008	0.005
R^2	0.877		0.984	
χ^2	1329.58***		7657.56***	
AIC	-15.174		-16.932	
N	186		128	

***: Significant at 0.01; **: Significant at 0.05

FIGURE 1

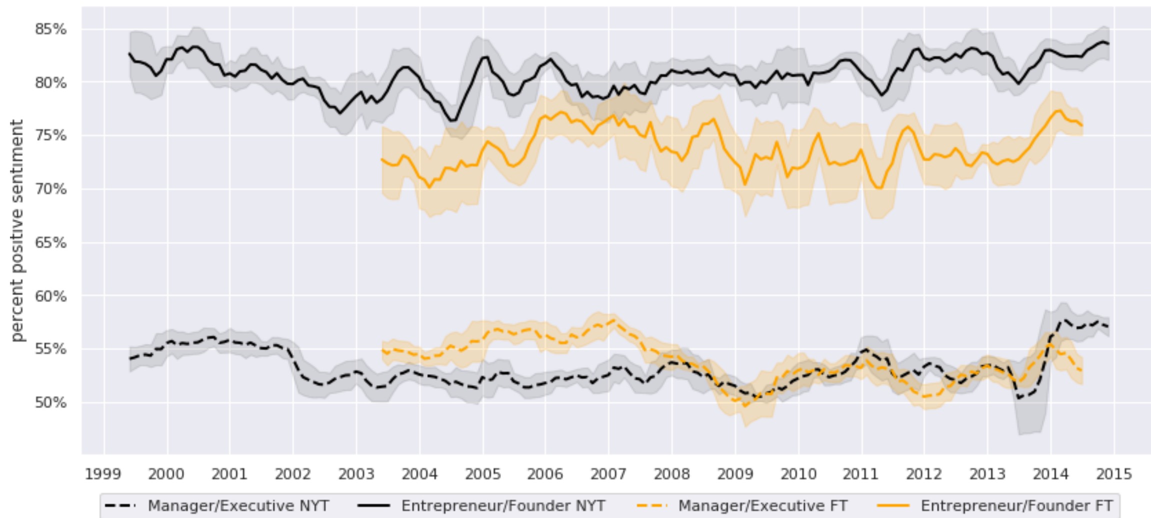
Monthly Frequency Count of Key Words



Notes. The lines (both continuous and dashed) represent the rolling mean of the observed data. We chose 6 months for the rolling window. That is, for every month, a 6-month window is taken and the average of the observed values contained in those 6 months is calculated. Then, the average is the new value for that month and the window rolls to the next month. The shaded areas represent the 95% confidence interval for the rolling mean. This interval contains the observed data with a 95% level of confidence, that is, the original observed value lies in the interval with a 95% probability.

FIGURE 2

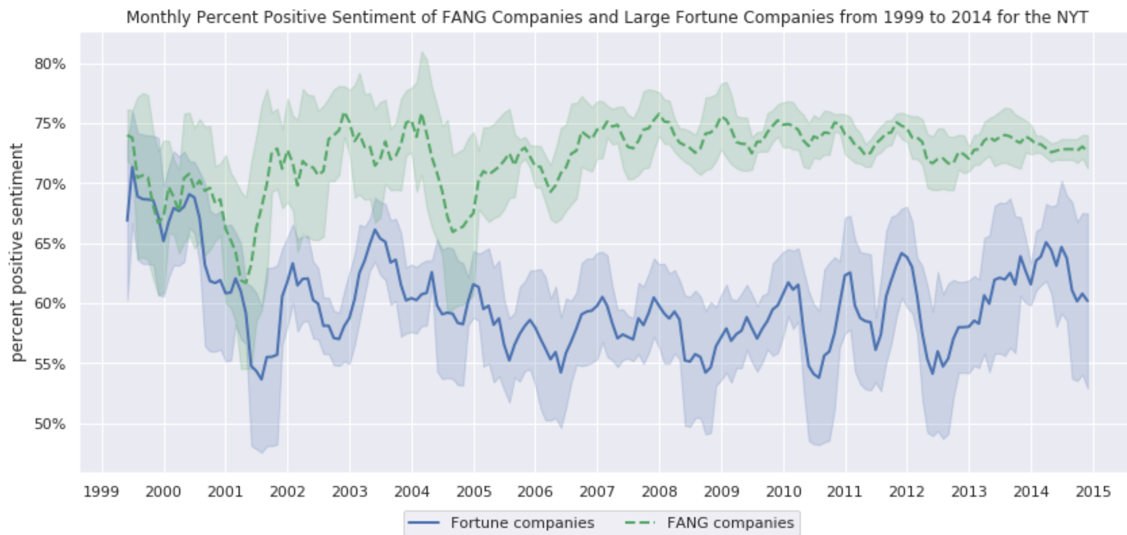
Monthly Frequency of Percent Positive Sentiments for
The New York Times and the Financial Times



Notes. For explanations of the lines and shaded areas, see notes to Table 1.

FIGURE 3

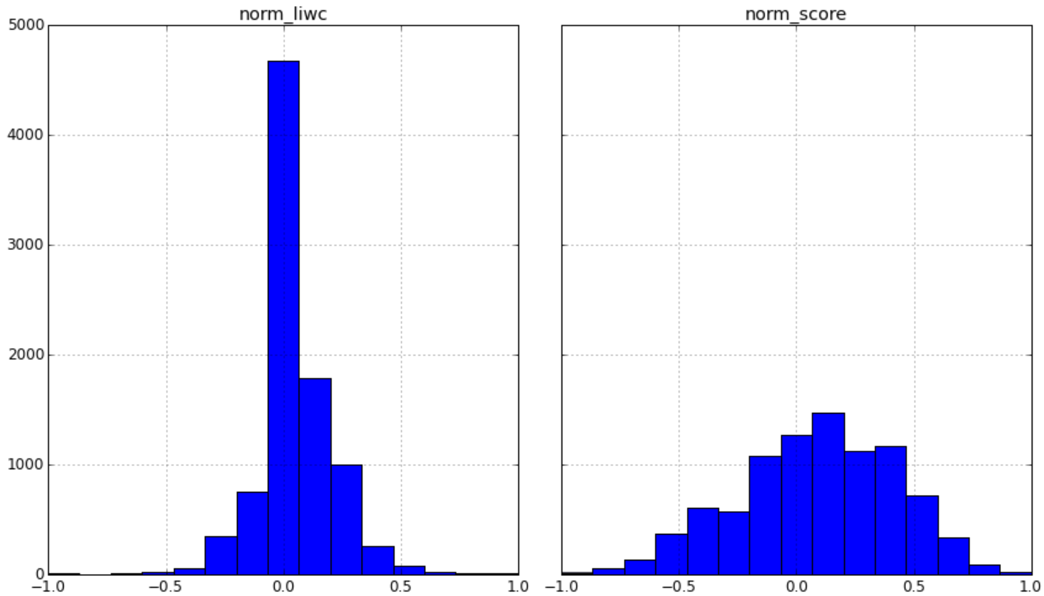
Monthly Percent Positive Sentiment of FANG Companies
and Large Fortune Companies from 1999 to 2014 for the NYT
(6 month moving average)



Notes. For explanations of the lines and shaded areas, see notes to Table 1.

FIGURE A1

Normalized Sentiment Score Histograms with 15 bins:
LIWC (left) vs. Purpose built (right)



REFERENCES

Andersson, M. & Koster, S. (2011) Sources of persistence in regional start-up rates – evidence from Sweden, ***Journal of Economic and Geography***, 11, 179-201.

Åstebro, T. & Chen, J. (2014). The entrepreneurial earnings puzzle: Mismeasurement or real?. ***Journal of Business Venturing***, *29*(1), 88-105.

Barnes, J., Klinger, R. & Schulte im Walde, S. (2017). Assessing state-of-the-art sentiment models on state-of-the-art sentiment datasets, **Proceedings of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis**. Copenhagen, Denmark September 7-11, 2-12.

Becker, G.S. & Murphy, K.M. (2000). ***Social economics: market behavior in a social environment***. Cambridge MA: Belknap Press.

Bernardo, A.E. & Welch, I. (2001). On the evolution of overconfidence and entrepreneurs, ***Journal of Economics & Management Strategy***, 10, 301-330.

Blackmore, S. (1999). ***The meme machine***. New York: Oxford University Press.

Blanchflower, D. G., Oswald, A., & Stutzer, A. (2001). Latent entrepreneurship across nations. ***European Economic Review***, *45*(4-6), 680-691.

Boczkowski, P. & Mitchelstein, E. (2013). ***The news gap: When the information preferences of the media and the public diverge***. Cambridge, MA: MIT Press.

Boettke, Peter J. Coyne, Christopher J. Context Matters: Institutions and Entrepreneurship in ***Foundations and Trends in Entrepreneurship*** 5:3 (2009).) Now: Hanover, MA.

Boyd, R., & Richerson, P.J. (2002) Group beneficial norms can spread rapidly in a structured population. ***Journal of Theoretical Biology***, 215(3), 287-296.

Brockhaus, R. H., & Horwitz, P. S. (1986). The psychology of the entrepreneur. ***Entrepreneurship: critical perspectives on business and management***, *2*, 260-283.

Campanale, C. (2010). Private equity returns in a model of entrepreneurial choice with learning, ***The B.E. Journal of Macroeconomics: Contributions***, 10(1), Article 22.

Carrington, W.J., McCue, K. and Pierce, B. (1996). 'The role of employer/employee interactions in labor market cycles: evidence from the self-employed'. ***Journal of Labor Economics,*** 14, 571-602.

Chiong, Raymond, Zongwen Fan, Zhongyi Hu, Marc T. P. Adam, Bernhard Lutz, and Dirk Neumann. 2018. "A Sentiment Analysis-Based Machine Learning Approach for Financial Market Prediction via News Disclosures." In *Proceedings of the Genetic and Evolutionary Computation Conference Companion*, 278–279. GECCO '18. New York, NY, USA: ACM. https://doi.org/10.1145/3205651.3205682.

Dawkins, R. (1989). ***The Selfish Gene***. New York: Oxford University Press.

De Meza, D. (2002). Overlending? ***Economic Journal***, 112, F17-F31.

Dosi, G. and Lovallo D. (1995) ***Rational Entrepreneurs or Optimistic Martyrs? Some Considerations on Technological Regimes, Corporate Entries, and the Evolutionary Role of Decision Biases***. IIASA Working Paper, WP-95-077.

Dridi, Amna, Mattia Atzeni, and Diego Reforgiato-Recupero. 2018. "FineNews: Fine-Grained Semantic Sentiment Analysis on Financial Microblogs and News." *International Journal of Machine Learning and Cybernetics*, March. https://doi.org/10.1007/s13042-018-0805-x.

Frankish, J.S., Roberts, R.G., Coad, A., Spears, T.C. & Storey, D.J. (2013). 'Do entrepreneurs really learn? Or do they just tell us they do?'. ***Industrial and Corporate Change***, 22(1), 73-106.

Gartner, W.B., K.G Shaver, N.M. Carter and P.D. Reynolds (2004). ***Handbook of Entrepreneurial Dynamics***. Thousand Oaks, CA: Sage Publications

George, G., Haas, M. R., & Pentland, A. (2014). Big data and management. ***Academy of Management Journal***, *57*(2), 321-326.

Granger, C.W. (1969). Investigating causal relationships by econometric models and cross-spectral methods, ***Econometrica***, *37*(3), 424-438.

Hall, R. E., & Woodward, S. E. (2010). The burden of the nondiversifiable risk of entrepreneurship. ***American Economic Review***, *100*(3), 1163-94.

Hamilton, B.H. (2000). 'Does entrepreneurship pay? An empirical analysis of the returns of self-employment'. ***The Journal of Political Economy***, 108(3), 604-631.

Headd, B. (2003). 'Redefining business success: distinguishing between closure and failure'. ***Small Business Economics***, 21, 51-61.

Heaton, J. & Lucas, D. (2000). Portfolio choice and asset prices: the importance of entrepreneurial risk, ***Journal of Finance***, 55, 1163—98.

Henley, A. (2004). Self-employment status: the role of state dependence and initial circumstances, ***Small Business Economics***, 22, 67-82.

Henrekson, M. & Sanandaji, T. (2014). Small business activity does not measure entrepreneurship, ***PNAS***, 111(5), 1760-1765.

Hyytinen, A., Ilmakunnas, P. & Toivanen, O. (2013). The return-to-entrepreneurship puzzle, ***Labour Economics***, 20, 57—67.

Kahn, J., Tobin, R., Massey, A., & Andersen, J. (2007) Measuring emotional expression with the Lingusitic Inquiry and Word Count. ***The American Journal of Psychology***, *120*(2), 263-286.

Kahneman, D. (2012). ***Thinking, Fast and Slow***. NY: Farrar, Strauss, Giroux

Kartashova, K. (2014). Private equity premium puzzle revisited. ***American Economic Review,*** *104*(10), 3297-3334.

Kihlstrom, R.E. & Laffont, J.J. (1979). A general equilibrium theory of firm formation based upon risk aversion, ***Journal of Political Economy***, 87, 719-49.

King, G., Schneer, B. & White, A. (2017). How the news media activate public expression and influence national agendas. ***Science*** *358*, 776-780.

Knight, F.H. (1921). ***Risk, Uncertainty and Profit.*** New York, NY: Houghton-Mifflin.

Loughran, T. & Mcdonald B. (2011). When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks. *The Journal of Finance* 66 (1): 35–65. https://doi.org/10.1111/j.1540-6261.2010.01625.x.

Loughran, T. & McDonald, B. (2016). Textual analysis in accounting and finance: A survey. Journal of Accounting Research 54 (4): 1187-1230.

Lovallo, D. & Kahneman, D. (2003). Delusions of success. How optimism undermines executives' decisions. ***Harvard Business Review***, 1-10.

McCombs, M. E., & Shaw, D. L. (1972). The agenda-setting function of mass media. ***Public Opinion Quarterly***, *36*(2), 176-187.

Moskowitz, T.J. & Vissing-Jørgensen, A. (2002). 'The returns to entrepreneurial investment: A private equity premium puzzle?'. ***American Economic Review***, 92, 745-778.

Nightingale, P. & Coad, A. (2014) Muppets and gazelles: political and methodological biases in entrepreneurship research. ***Industrial and Corporate Change,*** *23*(1), 113-143.

Nordhaus, W.D. (2004). ***Schumpetrian Profits in the American Economy: Theory and Measurement***. Working Paper No. 10433, NBER, Cambridge MA.

Parker, S.C. (2018), ***The Economics of Entrepreneurship, 2nd Edition***, Cambridge: Cambridge University Press.

Scott, W. Richard. (2014). ***Institutions and Organizations: Ideas, Interests and Identities (4th edition)***, Sage: Los Angeles.

Shane, S. (2009). Why encouraging more people to become entrepreneurs is bad public policy. ***Small Business Economics***, *33*(2), 141-149.

Slovic, P., Finucane, M. L., Peters, E., & MacGregor, D. G. (2002). The affect heuristic. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), ***Heuristics and biases: The***

psychology of intuitive judgment (pp. 397–420). New York: Cambridge University Press.

Torrance, W., Rauch, J., Aulet,W., Blum, L., Burker, B., D'Amrosio, T., de los Santos, K., Eesley, C., Green, W., Harrington, K., Jacquette, J., Kingma, B., Magelli, P., McConnell, G., Moore, D., Neeley, L., Song, M., Tan, T., Zoller, T., Zurbuchen, T. (2013). ***Entrepreneurship education comes of age on campus: The challenges of bringing entrepreneurship to higher education***. *Ewing Marion Kauffman Resarch Paper*, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2307987

Tversky, A. & Kahneman, D. Availability (1973): A heuristic for judging frequency and probability. ***Cognitive Psychology***, 4, 207-232.

Wanta, W., Golan, G., & Lee, C. (2004). Agenda setting and international news: Media influence on public perceptions of foreign nations. ***Journalism & Mass Communication Quarterly***, *81*(2), 364-377.

Wilson, Theresa, Janyce Wiebe, and Paul Hoffmann. 2005. "Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis." In *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing*, 347–354. HLT '05. Stroudsburg, PA, USA: Association for Computational Linguistics. https://doi.org/10.3115/1220575.1220619.

Wolpert, D. H. 1996. "The Lack of A Priori Distinctions Between Learning Algorithms." ***Neural Computation***, 8 (7): 1341 -90. https://doi.org/10.1162/neco.1996.8.7.1341.