Journal of Behavioral Finance

ISSN: 1542-7560 (Print) 1542-7579 (Online) Journal homepage: https://www.tandfonline.com/loi/hbhf20

Soft Strategic Information and IPO Underpricing

James C. Brau, James Cicon & Grant McQueen

To cite this article: James C. Brau, James Cicon & Grant McQueen (2016) Soft Strategic Information and IPO Underpricing, Journal of Behavioral Finance, 17:1, 1-17, DOI: 10.1080/15427560.2016.1133619

To link to this article: https://doi.org/10.1080/15427560.2016.1133619

JOURNAL OF BEHAVIORAL FINANCE
2016, VOL. 17, NO. 1, 1–17
http://dx.doi.org/10.1080/15427560.2016.1133619

Soft Strategic Information and IPO Underpricing

James C. Brau[a], James Cicon[b], and Grant McQueen[a]

[a]Brigham Young University; [b]University of Central Missouri

ABSTRACT

Using content analysis, we measure the impact of soft information, derived from words in initial public offering (IPO) registration documents, on IPO pricing efficiency. First, using 2,298 U.S. IPOs from 1996–2008, we find that an IPO document's strategic tone correlates positively with the stock's first-day return; more frequent usage of positive and/or less frequent usage of negative strategic words leads to more IPO underpricing. Second, we find that an IPO document's strategic tone is negatively correlated with the stock's long-run return. Together, these findings imply that investors initially misprice soft information in registration statements, which mispricing is eventually corrected. Additionally, we create new content-analysis libraries for strategic words and introduce a survey-based library creation method and word-weighting system.

KEYWORDS
Soft information; IPOs; Textual analysis; Underpricing; Word counts; Content analysis

Introduction

Researchers have worked to explain why initial public offerings (IPOs) are underpriced, on average, since the seminal work of Stoll and Curley [1970], Logue [1973], and Ibbotson [1975] documented the anomaly. Theoretical models, such as Baron [1982] and Rock [1986], began explaining IPO underpricing in the context of asymmetric information. Empirical documentation of, and explanation for, the anomaly has focused on quantifiable market and financial statement data (e.g., Beatty and Ritter [1986], Hanley [1993]). Most relevant to our research is Bradley and Jordan's [2002] finding that publicly available quantitative information (such as the share overhang and file range amendments) can predict a stock's first-day return. Along with this quantitative data, however, is underutilized "soft information"—the words themselves—found in IPO registration statements. We use content analysis to examine whether the market for IPOs is efficient with respect to publicly available soft strategic information.

Content analysis quantifies semantic information to make "replicable and valid inferences from texts to the contexts of their use" (Krippendorf [2004]). Table 1 summarizes early business and recent finance research linking soft information in business documents (e.g., earnings announcement or annual report) to an economic outcome (e.g., event return or lawsuit).[1] Brockman and Cicon [2013], for example, found that the text has a greater influence than the numbers on the economic outcome. Whereas content analysis is the methodology for studying the content of all forms of communication (e.g., text and speech), textual analysis focuses on just text. We analyze text but use the broader term content analysis to describe our methodology. Specifically, we quantify a registration statement's strategic tone using the frequency of words associated with positive and negative business strategies. For example, words such as "innovation" or "differentiated" are associated with positive business strategies, and words such as "stagnant" and "commodity" are associated with negative competitive situations. We test whether the registration document's strategic tone is correlated with IPO underpricing, the percent difference between the offer price and the price at the end of the first trading day.

We choose to study strategic words, rather than finance words, for two reasons. First, finance words are more likely to be redundant with the well-studied finance numbers (i.e., accounting statements). For example, a written explanation of profits is not only redundant to the measures of profit in the accompanying accounting statements but is also less precise. Writing that "profits are high" will have little value influence beyond the fact that net income is, say, $4,967,000 together with exact levels of sales, assets, and equity for the current and prior years. Second, at the IPO stage, a strong strategy is consistently value-enhancing whereas a financial tone's effect on value can vary. For example, in some industries (e.g., technology) and in some periods (e.g., the late

CONTACT Grant McQueen mcqueen@byu.edu William Edward Professor of Finance, Marriott School of Management, Brigham Young University, 636 TNRB, Provo, UT 84602.

Table 1. Articles using content analysis (soft information) to explain financial outcomes.

Article	Source of Information/Content	Impacted Variable(s)
Bettman and Weitz [1983]	letters to shareholders/degree of self-serving attribution	corporate performance
Antweiler and Frank [2004]	internet message board/bullishness	stock returns and volume
Li [2006]	annual report/risk sentiment	future earnings and stock returns
Tetlock [2007]	media/pessimism	stock returns and volume
Demers and Vega [2008]	earnings announcement/optimism	announcement returns
Engelberg [2008]	earnings announcement/negativity	announcement stock price drift
Davis, Piger, and Sedor [2008]	firm discloser/tone	cost of capital
Kothari, Li, and Short [2008]	firm discloser/tone	cost of capital
Tetlock, Sarr-Tsechansky, and Macskassy [2008]	news story/negativism	earnings and returns
Loughran and McDonald [2010]	annual report/financial tone	returns
Hanley and Hoberg [2010]	IPO document/uniqueness	mispricing
Kogan, Routledge, Sagi, and Smith [2010]	management disclosure and analysis in annual report	share liquidity
Arnold, Fishe, and North [2010]	IP document/ambiguity	returns
Rogers, Van Burkirk, and Zechman [2011]	earnings announcement/optimism	litigations
Baginski, Demers, Wang, and Yu [2011]	management earnings forecasts/sentiment and certainty	announcement returns
Jegadeesh and Wu [2011]	annual report tone	returns
Hanley and Hoberg [2012]	IPO document/uniqueness	litigations
Garcia [2012]	*NYT*'s financial news sentiment	asset prices
Gurun and Butler [2012]	news reports/negativism	equity valuations
Price, Doran, Peterson, and Bliss [2012]	quarterly conference call/tone	returns
Brockman and Cicon [2013]	management earnings forecast/optimism and certainty	announcement returns
Ferris, Hao, and Liao [2013]	IPO document/conservatism	returns
Loughran and McDonald [2013]	IPO document/uncertainty	returns

1990s), firms with weak profits actually received high valuation multiples. Such appears to be the case since Loughran and McDonald [2011] find that negative financial words in an annual report hurt the value of an established firm but Ferris, Hao, and Liao [2013] find that negative financial words in an IPO prospectus help the value of a technology firm going public.

We apply content analysis to IPOs and make three contributions. The first two address the economic issue of market efficiency and the third address the creation of word lists.

First, we show that the relative frequency of positive and negative strategic words in the Securities and Exchange Commission (SEC) Form S-1 (IPO registration document/prospectus) correlate statistically with the first-day return (i.e., IPO underpricing) after controlling for factors known to be associated with the first-day return, and we show that this correlation is economically meaningful. A higher frequency of positive strategic words or a lower frequency of negative strategic words is associated with a larger first-day return. Thus, we contribute to the growing evidence of a new inefficiency: soft information in the registration statement is predictably correlated with the degree of IPO underpricing.

This first contribution extends the work of Hanley and Hoberg [2010], Arnold et al. [2010], Ferris et al. [2013], and Loughran and McDonald [2013], who apply content analysis to IPOs. We differ from these three studies in both measure and purpose. Hanley and Hoberg [2010] measure an IPO registration's informative content (word usage uniqueness). They use informative content as a proxy for the underwriter's effort and the document's disclosure and show that more effort and disclosure lead to more efficient initial prices. Arnold et al. [2010] measure an IPO document's ambiguity as the ratio of the Risk Factor section's word count to the total prospectus's word count. They find a correlation between the ratio and underpricing, which suggests investors require greater IPO underpricing as compensation for ambiguity. Ferris et al. [2013] measure the frequency of negative financial words in the prospectus. They use this frequency as a conservatism proxy and link it to greater mispricing (high first-day return). Loughran and McDonald [2013] document a positive correlation between uncertain text in the IPO and the first-day return. Unlike these four papers, we do not measure uniqueness, ambiguity, financial tone, or the degree of uncertainty. We measure an IPO registration's strategic tone—words associated with positive and negative business strategies and show that the more positive the strategic tone, the greater the degree of underpricing.

The evidence in these four papers and our evidence that strategic tone correlates with the initial return indicate that either the initial price or the price at the end of the first day is inefficient. However, this evidence from the first day of trading is silent on whether the IPO price or the price at the end of the first day is wrong. In our second contribution, we test whether investment bankers' or investors' use of soft information is to blame for the IPO first-day pricing inefficiency. That is, we test whether the offer price under-adjusts or the closing price over-adjusts to strategy words in the prospectus, given that this soft information is widely available. The correlation between the registration document's strategic tone and the stock's first-day return can be driven by contrasting explanations: (1) IPO underwriters could correctly incorporate soft information into the IPO price, with overly optimistic or pessimistic investors subsequently setting an incorrect price as they are swayed by strategic words, or (2) the underwriters themselves could ignore or undervalue some soft strategic information that savvy investors quickly incorporate into the price during the first day of trading. Our evidence from long-run IPO returns is consistent with the former: investors appear to initially misprice this soft information, adding to the inefficiency of the IPO process.

Our first two contributions can be summarized as follows: soft strategic information in the registration statement is priced, albeit initially mispriced, in the secondary market. That is, with respect to soft strategic information in the prospectus, the anomaly is not initial underpricing of the offer price by investment bankers but rather subsequent overpricing by investors.

The third contribution is similar to that of Loughran and McDonald [2011], who created finance-related word libraries. We contribute two new strategy-related word libraries. Additionally, we contribute a novel survey-based approach to creating libraries. The two most common sets of word lists used in textual analysis are the Harvard Psychosociological Dictionary and DICTION, a popular textual analysis software program. Both lists have been previously used in financial-economics research. For example, Das and Chen [2007] use the Harvard Dictionary to measure sentiment in messages on stock message boards. Tetlock [2007] links pessimistic words in the *Wall Street Journal*'s "Abreast of the Market" column with stock returns and uses the Harvard Dictionary's word categories to create a measure of "pessimism." Ober, Zhao, Davis, and Alexander [1999] use DICTION's lists to measure 10-K filings' degree of "certainty" and find that language of certainty (compared to hedges and less-direct words) is not correlated with profitability. Loughran and McDonald [2011] find that these standard dictionaries are not specialized enough to measure the tone of business-related documents. On page 36 of their paper, they state that "…almost three-fourths (73.8%) of the negative word counts according to the Harvard list are attributable to words that are typically not negative in a financial context." Furthermore, for our purposes, we need words associated with good and bad strategic positions; however, words such as "branded," "copyright," and "differentiated," are not in either standard word list and words such as "barrier" are categorized in negative lists despite the typically positive connotation (i.e., barrier to entry) in a strategy discussion. Consequently, we create our own lists of strategy words.

Our survey-based library creation method has three advantages: (a) relative to researcher-created lists, we mitigate researcher bias or word mining; (b) relative to standard lists (e.g., Harvard Psychosociological Dictionary), we have specific context since the words are rated by business people who rate with respect to business strategy; (c) relative to both, we can create a continuous measure of a word's tone rather than dichotomous categories of positive and negative sounding strategy words.[2]

In addition to these three contributions, we consider several possible improvements to the content-analysis methodology used in the finance literature. We identify necessary conditions for two measures of content to be reduced to one dimension (e.g., subtracting measures of "pessimism" from "optimism") to create only one score. We introduce a word-weighting system based on survey responses. Prior researchers choose one of many weighting systems. Our system eliminates any searching-related researcher bias by allowing disinterested professionals to determine the proper weights. We also expand and measure the use of negation words and consider symbols, such as trademark (™), in addition to words.

Data

IPOs and S-1s

Our initial list of 4,509 firm-commitment IPOs between January 1996 and December 2008, inclusive, comes from Securities Data Corporation's (SDC) new issues database. We end the sample of IPOs in 2008 because in our long-run tests we require three years of post-IPO abnormal returns. Thus, our return data run through December 2011. We delete 523 non-U.S. issuers. Following Hanley and Hoberg [2010, 2012], we exclude 981 issuers in the finance industry as identified by SDC. This process eliminates American Depositary Receipts (ADRs), real estate investment trusts (REITs), and closed-end mutual funds from our sample. We search the SEC's online repository (EDGAR) to match (using Central Index Keys, or CIKs) each IPO company to its electronic version of the registration statement (S-1 or S-1/A) nearest to, but before, the IPO date. We match 2,367 IPOs. We delete 69 more observations because they could not be located on CRSP, do not have a closing price on the first trading day, or appear twice in the SDC database. Tables 2 and 3 report summary statistics for all remaining 2,298 observations. However, our regression results are based on 2,247 total observations with complete data on all independent variables from EDGAR, SDC, CRSP, Compustat, and Jay Ritter's data web page. We retrieve the registration documents from EDGAR; the original list of IPOs, IPO proceeds, offer price, pre-offer price change, number of amendments, secondary shares offered, use of proceeds, underwriter name, underwriter spread, VC status, and auditor name from SDC; the first-day closing price and

Table 2. IPO sample distribution frequencies over time and across industries.

Panel A. Distribution by Issue Year

Issue Year	Frequency	Percent	Cumulative Frequency	Cumulative Percent	Initial Return Mean	Initial Return Median	Strategic Tone Mean
1996	330	14.36	330	14.4	13.5%	8.7%	−0.005
1997	344	14.97	674	29.3	14.2%	9.4%	−0.094
1998	211	9.18	885	38.5	23.6%	10.0%	0.041
1999	386	16.8	1271	55.3	76.8%	44.3%	0.400
2000	312	13.58	1583	68.9	59.3%	29.6%	0.439
2001	64	2.79	1647	71.7	14.7%	10.2%	−0.126
2002	59	2.57	1706	74.2	7.3%	6.3%	−0.518
2003	45	1.96	1751	76.2	13.9%	10.3%	−0.325
2004	138	6.01	1889	82.2	12.8%	8.4%	−0.282
2005	127	5.53	2016	87.7	10.0%	6.5%	−0.479
2006	131	5.7	2147	93.4	11.1%	5.0%	−0.562
2007	133	5.79	2280	99.2	14.3%	6.9%	−0.192
2008	18	0.78	2298	100	6.2%	−1.5%	−0.438

Panel B. Distribution by Industry

Industry	Frequency	Percent	Cumulative Frequency	Cumulative Percent	Strategic Tone Mean
Agriculture, Forestry, and Fishing	7	0.30	7	0.30	−0.571
Mining	83	3.61	90	3.9	−1.605
Construction	22	0.96	112	4.9	−0.662
Manufacturing	813	35.38	925	40.3	0.024
Transportation, Communications, Electric, Gas, and Sanitary Services	226	9.83	1151	50.1	−0.535
Wholesale Trade	64	2.79	1215	52.9	−0.339
Retail Trade	155	6.74	1370	59.6	−0.250
Services	928	40.38	2298	100	0.344

The IPO sample is drawn from Securities Data Company's (SDC) New Issues Database and matched with the Securities and Exchange Commission's registrations statement from EDGAR. We exclude non-U.S. issuers and issuers in the finance industry. Industries are based on Kahle and Walking [1996]. Strategic Tone Mean is the average *Strategic Tone* of the IPOs based on normalized, unweighted strategic word frequency as defined in Equations 1–3.

three-week prior market return from CRSP; and the book value of equity from Compustat. We also use the Internet dummy status, underwriter rank, firm founding date, dual-class status, number of IPOs in the prior three months, and average prior three-month underpricing from Jay Ritter's web page.

Table 2 reports the distribution of IPOs over time and industry. The drop in IPOs for 2001 and 2008 correspond with the stock market declines often referred to as the end of the dot-com and real-estate bubbles, respectively. Our two-digit SIC classifications are based upon Kahle and Walkling [1996]. The most represented industries in our sample are Services (40.38%) and Manufacturing (35.38%). The mean strategic tone data (last column) is defined later in this paper.

Word lists

Neither the Harvard Psychosociological Dictionary nor DICTION's lists are specific enough for our study of business strategy; therefore, we create our own. We create these lists in two steps: (1) compiling a list of strategy-related words and (2) rating the words' strategic tone by current MBA students and recent MBA graduates. The first step introduces potential biases as we choose which words to include on the initial list. However, in the second step, any biases are mitigated by survey participants who ultimately decide on which words are positive and negative.

In step one, we gather strategy-related words from the subject index of Hitt, Ireland, and Hoskisson's [2005] *Strategic Management* and words in DICTION that could describe a business strategy. We then exclude words from Loughran and McDonald's [2011] list of finance and legal words (labeled *Fin-Neg, Fin-Pos*, and *Fin-Lit* lists in their paper). Thus, our list of strategy-related words does not overlap with existing lists of finance and legal words. This process generates a list of 2,349 words in 555 related-word groups. For example, the six words "brand," "branded," "brander," "branders," "branding," and "brands" form one word group. We manually group words together based on typical business usage. We do not use Porter's [1980] stemming algorithm to form word groups since, for example, it links "competition" and "competitive" with the same word stem "competit." When survey participants were able to rate these two words separately, they considered "competition" a negative word and "competitive" a positive word. Hence, we keep these words separate, whereas stemming would group them together.

In step two we surveyed 399 individuals asking them to rate the step-one words on a 5-point Likert-type response scale, with 1 indicating a very poor business strategy and 5 indicating a very good business strategy. We received 305 completed surveys, a 76% response rate. About 40% of the participants were recent MBA graduates (average age 31 and average GMAT 672); about 40% were MBA students between their first and second year of the program (average age 30 and average GMAT 675); and the roughly 20% remaining were Executive MBA students almost finished with their degree (average age 39). All participants had completed at least one finance and one strategy class. Participation was voluntary; students had the option of entering their name at the end of the survey to enter a raffle for an iPod Touch 32G (approximate value of $300).

Survey participants were sent an email stating, "I am researching whether strategic words in an IPO prospectus influence how much investors pay for the newly traded stock. I need your help in rating strategic words. The survey takes about 10 minutes." The participants then could click on a link and take the online questionnaire. Note that the context is specific, strategy words in an IPO. Participants rated only one word from each word group; we assumed that the other words in the group would garner a similar rating. For example, if the average student rating of the word "loyal" was 4.15, we assigned this same rating to the un-rated words "loyally," "loyalness," "loyalties," and "loyalty." To increase response rates, we asked participants to rate only 150 words randomly selected from the full word list. As a result, the words were rated by different numbers of participants. Words received an average of 43 rankings. The appendix gives the highest and lowest ten words by rating.

Our second library-creation step has three advantages: (a) the researchers do not categorize or rate words themselves, eliminating the possibility of data mining; (b) the library incorporates the views of many business professionals (rather than a few academics) in a specific business-strategy context; and (c) the words are on a continuum from 1 (bad) to 5 (good), rather than a discrete scale (indicator variables).

When creating our word lists, we control for prefixes and negation. Prefix control allows for "ability" and "approve" to be on the positive word list as well as "*in*ability" and "*dis*approve" on the negative word list. A word is negated when preceded by a word such as "not," that reverses the word's meaning. We incorporate Loughran and McDonald's [2011] six negation words: "neither," "never," "no," "nobody," "none," and "not." To their list we add the following 22 negation words: "aren't," "cannot," "can't," "couldn't," "didn't," "doesn't," "don't," "hadn't," "hasn't," "haven't," "isn't," "mustn't," "needn't," "nor," "nothing," "nowhere," "shouldn't," "wasn't," "weren't," "without," "won't," and "wouldn't." Additionally, we add "too" to our negation list since, for example, "too strong" often carries a negative connotation whereas "strong" is generally positive. Thus, in the phrase "was always right," we consider "right" as a positive word, whereas, in the phrase "was never right," we consider "right" combined with "never" as negative. We account for negation in both negative and positive word counts. Negation of negative words occurs frequently in registration statements. We negate when any one of these twenty-nine negation words occurs within two words preceding a strategy word (e.g., "not actually reduced" would also be considered positive).

Strategic Tone

We define an IPO registration document's *Strategic Tone* as the difference between the positive and negative, peer-normalized word frequencies based on the positive and negative strategic word lists as follows:

$$Posfreq_j = \frac{\sum_{i}^{N^+} Positive\_count_{i,j}}{Total\_count_j} \quad \& \quad Negfreq_j = \frac{\sum_{i}^{N^-} Negative\_count_{i,j}}{Total\_count_j}, \tag{1}$$

$$Positive\ Score_j = \frac{Posfreq_j - \mu_{Posfreq}}{\sigma_{Posfreq}} \quad \& \quad Negative\ Score_j = \frac{Negfreq_j - \mu_{Negfreq}}{\sigma_{Negfreq}}, \tag{2}$$

$$Strategic\ Tone_j = Positive\ Score_j - Negative\ Score_j, \tag{3}$$

where $Positive\_count_{i,j}$ is the number of times word i occurs in document j; $Total\_count_j$ is the total number of words in document j; N^+ is the number of words in the positive word list; $Posfreq_j$ is the frequency of positive words in document j; $\mu_{Posfreq}$ is the average *Posfreq* across all 2,298 documents in our sample; and $\sigma_{Posfreq}$ is the standard deviation of *Posfreq* across our documents. Similar definitions apply for negative words. Words not on our strategic lists are not counted in the numerators of Equation 1 but are counted in the denominator that controls for document length. Each listed-word occurrence initially gets equal weight. We include words with an average survey rating above 3.0 on the positive list and words with an average survey rating 3.0 or below on the negative list (scale ranges from 1 to 5).

Equation 1, simple frequency, is the measure used by Davis, Piger, and Sedor [2008], Rogers, Van Buskirk, and

Zechman [2009], and Loughran and McDonald [2011, 2013]. The standardization in Equation 2 transforms the frequency measure in (1) into a z-score for ease of interpretation, comparison, and combination. For purposes of interpretation, reporting a document's *Posfreq* score as 1.6% is not as informative as saying the document's *Posfreq* is one standard deviation above the mean. Regarding comparison, positive words are more common than negative words in our documents, so a *Posfreq* of 0.6% does not mean the same thing as a *Negfreq* of 0.6%. Regarding combination, standardization facilitates the collapsing of two dimensions (i.e., positive and negative scores) into one (i.e., *Strategic Tone*) in Equation 3.

Typically, content analysis standardizes word frequencies (i.e., Equation 2) relative to the frequency in a sample of texts of political speeches, newspaper editorials, business reports, scientific documents, etc. (e.g., Hart [1984]). In contrast, we standardize frequencies relative to all other registration documents in our sample. We peer standardize because Cicon, Clarke, Ferris, and Jayaraman [2010] find that word frequencies in business-related filings are different than those in a broad cross-section of English text. For example, EDGAR filings are written with less "optimism" but more "certainty" than other types of documents. Consequently, standardizing our documents relative to a broad collection of English text would result in scores having non-zero means and standard deviations greater than one.

Equation 3 collapses the two dimensions (i.e., positive and negative scores) down into one dimension (i.e., *Strategic Tone*). We document the legitimacy of this step later when we report the empirical results. The number of semantic dimensions is often ignored in content analysis since an additive statistical model (no cross products) is assumed and coefficient restrictions are typically made without testing. For example, content analysis is often implemented using DICTION software (Hart [2000, 2001]). At the core of DICTION are thirty-three non-overlapping lists of words that capture semantic features used in English communication (West [2001]). For example, Davis et al. [2008] measure the optimism and pessimism of earnings press releases using the following six DICTION lists: *optimism* = *praise* + *satisfaction* + *inspiration*, with *pessimism* = *blame* + *hardship* + *denial*. Mapping a six-dimensional space down to one or two dimensions can cause three types of problems. First, the additive model can be misspecified. For example, *satisfaction*'s influence may be conditional on the degree of *inspiration* (significant cross-product coefficient) causing the coefficient on *inspiration* (from a model ignoring the cross-product) to be biased. Second, a single component may be insignificant even if the sum of the components is statistically significant. For example, *denial* may add nothing to *pessimism*'s predictive ability (*denial*'s correlation with event returns is zero), with *pessimism*'s significance (joint coefficient) being driven only by the words in *blame* and *hardship*. Third, the combination of scores (decrease in dimensions) can be incomplete. For example, the assumed separate dimensions of *optimism* and *pessimism* may actually collapse to one dimension (the slope of *optimism* equals the negative of the slope of *pessimism*) we will call "Net-*mism*."[3] We specifically test for dimension. In our subsequent regressions, the lack of a significant interactive term and the approximate symmetry of the positive and negative coefficients in our Model 1 indicate that it is appropriate to subtract the negative from the positive score to create overall tone in Equation 3.

To take advantage of the spectrum of word rankings, we also consider a modified version of Equation 1 after weighting words. In our base case, each positive and negative word in Equation 1 is counted as one, even though, for example, "leader," "faster," and "centralization" are not equally positive (survey rankings of 4.45, 4.00, and 3.03, respectively). Our weights allow words to count between zero and two depending on the strength of the survey ranking. For positive words, we subtract three from the average rank, so, for example "leader" counts as 1.45, "faster" as 1.00, and "centralization" as 0.03 in the numerator of Equation 1. We weight negative words in a similar fashion, specifically subtracting the average survey rank from 3 that allows us to create the weighted score.

Our weighting system, like that proposed in Jegadeesh and Wu [2011], takes the weighting system out of the hands of the researcher thus limiting the possibility of searching over multiple weighting functions. They document cases where weighting systems borrowed from the document retrieval literature can lead to incorrect conclusions. They develop a new weighting system, based on market reaction to words out of sample.[4]

The first row in Table 3 reports that the average word in our list received a score of 3.01 on a scale of 1 ("Very Poor") to 5 ("Very Good"). The lowest rating, 1.22, was for "bribe"; the highest rating, 4.46, was for "exemplary." The next two rows document that positive words are a lot more common than negative words in prospectuses.

For comparison purposes, we test the correlation between IPO first-day returns and normalized scores using Loughran and McDonald's [2011] positive (*Fin-Positive*) and negative (*Fin-Negative*) finance word lists that were drawn from a sample of 10-Ks and customized for finance applications. Panel B of Table 3 reports correlation coefficients for normalized scores using our strategy word list and Loughran and McDonald's [2011] two finance word lists. Both *Strategic Tone* and *Fin-Negative* are significantly correlated with *First-day Return*. Panel B also reports significant but small correlations between a document's tone and its Loughran and McDonald's

Table 3. Summary statistics for registration statement frequencies and scores of strategic tone.

Panel A. Descriptive Statistics

Variable	Mean	Standard Deviation	Minimum	Maximum
Word Survey Rating	3.01	0.803	1.22	4.46
Negative Count	1,850	398	610	4,034
Positive Count	9,435	1,642	2,479	15,063

Panel B. Correlation Table

	Under-pricing	*Strategic Tone*	*L&M Fin-Negative*
Strategic Tone	0.24005		
	<.0001		
L&M Fin-Negative	0.12313	0.06144	
	<.0001	0.0032	
L&M Fin-Positive	−0.00312	0.26975	−0.03497
	0.8813	<.0001	0.0937

The strategy scores are based on our two strategy word lists. The L&M scores are based on Loughran and McDonald's [2011] finance word lists. The summary statistics are for the 2,247 IPO registration documents from 1996 to 2008 found in both the SDC and EDGAR data bases and with independent variables available in CRSP and Compustat.

[2011] finance scores even though we exclude the finance words from our strategy library. Consequently, in our multivariate tests we consider the effect of finance and strategy scores on mispricing separately as well as jointly.

Strategic tone across time, industry, and underwriter

The last columns in Panels A and B of Table 2 indicate that *Strategic Tone* is significantly higher than average (unreported p-values less than 0.01) in 1999 and 2000 and in the services industry that includes relatively more technology-related firms. *Strategic Tone* is significantly lower than average (unreported p-values less than 0.01) in 2002 and 2004–2006, and in the mining, construction, and transportation-communication-utilities industries. Thus, the tone of the prospectus appears to be a contributing factor to the frothy, over-valued tech IPOs in the late 1990s (see Loughran and Ritter [2004]).

Table 4 reports the mean and rank of strategic tone and adjusted-strategic tone by underwriter. To manage the size of the table, we limit our reporting to firms that underwrote ten or more IPOs in our sample. Adjusted strategic tone is the *Strategic Tone*, after controlling for the year and industry fixed effects evident in Table 2, along with a control for Internet-related IPOs. Specifically, adjusted strategic tone is the strategic tone less the mean strategic tone for IPOs in the same industry and year. The difference between unadjusted and adjusted tone can be illustrated using Fleet Boston's rankings, reported in the last two columns. Looking only at prospectuses, the underwriter is unusually positive (ranked eighth) when describing the strategies of the firms it takes public. However, after considering the number of technology firms and firms going public in the frothy years, this apparent tendency to "hype" the strategy disappears. The apparent eighth-ranked firm drops to number 34 (out of 43 total firms) after considering fixed effects of industry and year.

Dependent and independent variables

To test whether registration documents' strategic soft information influences underpricing, we create the following dependent and independent variables that control for extant determinants of underpricing found in the literature.

Dependent variable

First-day Return (Underpricing) = IPO initial return, the percentage change in price from the IPO price to closing price on the first day of trading.

Offering characteristics

IPO Proceeds = Gross proceeds of the IPO (total shares offered times offer price), measured in millions of 2008 purchasing-power dollars (Habib and Ljungqvist [1998]).

Pre-offer Price Change = Partial stock price adjustment, percentage difference between the mid-range filing price in the original S-1 and the final offer price (Hanley [1993]).

Underwriter Spread = Percent of total proceeds going to the underwriting syndicate (Chen and Ritter [2000]).

Refinance Debt Dummy = Binary variable indicating that one of the proceeds' uses is to service debt (Beatty and Ritter [1986]).

Number of Amendments = The number of times the S-1 document was amended (Ang and Brau [2003]).

Secondary Shares Offered = The percent of secondary shares in the IPO (Leland and Pyle [1977]).

Dual Share Class Dummy = Binary variable indicating more than one class of common stock (Smart and Zutter [2003]).

Table 4. Mean, median, and ranking of adjusted strategic tone by active IPO underwriters.

Underwriter	Count	Adjusted Tone Mean	Unadjusted Tone Mean	*p*-value	Rank after Adjustments	Unadjusted Rank
Thomas Weisel Partners	19	0.462	0.539	0.010	1	3
BancAmerica Robertson Stephens	19	0.334	0.522	0.124	2	4
Piper Jaffray	17	0.316	0.159	0.123	3	14
Credit Suisse	31	0.281	0.009	0.115	4	24
UBS Warburg	15	0.278	0.278	0.187	5	12
Banc of America Securities	37	0.274	0.064	0.028	6	21
Deutsche Bank	20	0.251	−0.009	0.284	7	25
Hambrecht & Quist	44	0.246	0.501	0.017	8	5
SG Cowen Securities	23	0.226	0.294	0.258	9	10
Deutsche Banc Alex Brown	32	0.213	0.613	0.209	10	2
Morgan Stanley	78	0.207	−0.031	0.070	11	27
US Bancorp Piper Jaffray	14	0.182	0.421	0.300	12	7
Wachovia Securities	10	0.164	−0.514	0.648	13	39
CIBC World Markets	18	0.160	0.293	0.356	14	11
William Blair	15	0.147	0.114	0.573	15	18
BancBoston Robertson Stephens	45	0.098	0.659	0.413	16	1
Cowen	14	0.094	0.040	0.575	17	22
JP Morgan	59	0.080	−0.097	0.450	18	28
NationsBanc Montgomery Securities	19	0.074	0.220	0.727	19	13
Merrill Lynch	115	0.073	0.022	0.353	20	23
Bear Stearns	54	0.027	0.136	0.825	21	16
Morgan Stanley Dean Witter	99	0.016	0.332	0.856	22	9
Friedman Billings Ramsey Group	13	−0.018	−0.191	0.969	23	33
Goldman Sachs	196	−0.027	−0.012	0.725	24	26
Cruttenden Roth	14	−0.030	0.117	0.849	25	17
Donaldson Lufkin & Jenrette	76	−0.035	0.153	0.703	26	15
BT Alex Brown	32	−0.045	0.101	0.775	27	20
Chase H&Q	23	−0.050	0.460	0.819	28	6
WR Hambrecht	15	−0.057	−0.169	0.790	29	32
Credit Suisse First Boston	155	−0.066	0.108	0.394	30	19
Jefferies	14	−0.122	−0.307	0.383	31	35
Raymond James & Associates	11	−0.143	−0.676	0.738	32	40
Lehman Brothers	112	−0.154	−0.360	0.154	33	36
Fleet Boston Corp	24	−0.181	0.392	0.299	34	8
Alex Brown & Sons	11	−0.189	−0.295	0.358	35	34
Smith Barney	13	−0.205	−0.167	0.621	36	31
Salomon Smith Barney	41	−0.205	−0.131	0.273	37	29
Needham	10	−0.217	−0.139	0.421	38	30
Montgomery Securities	17	−0.293	−0.361	0.633	39	37
Citigroup	37	−0.393	−0.843	0.060	40	41
Prudential Securities	20	−0.429	−0.435	0.254	41	38
UBS Securities	37	−0.440	−1.016	0.014	42	43
AG Edwards & Sons	19	−0.464	−0.995	0.104	43	42

Count is the number of IPOs in our 1996–2008 sample attributed to the underwriter. Only firms underwriting 10 or more IPOs are reported. Unadjusted tone is the normalized *Strategic Tone* defined in Equations 1–3. Adjusted tone is the strategic tone after controlling for year, internet, and industry fixed effects. Ranks are based on sorting by adjusted or unadjusted mean strategic tone of the firm. The *p*-value indicates whether the mean Adjusted Tone is significantly different from zero.

High Tech Dummy = Binary variable indicating a high-tech firm (Field and Hanka [2001]).

Internet IPO Dummy = Binary variable indicating an Internet-based firm (Loughran and Ritter [2004]).

Third-party certification

Underwriter Prestige = Lead underwriter rank for year prior to IPO (Booth and Smith [1986], Carter and Manaster [1990], Beatty and Ritter [1986], Carter, Dark and Singh [1998]. Ranges from 0 (worst) to 9 (best)).

Venture Capital Backing = Binary variable indicating the presence of VC backing prior to the IPO (Megginson and Weiss [1991], Barry, Muscarella, Peavy and Vetsuypens [1990], Brav and Gompers [1997]).

Big-Six Auditor Dummy = Binary variable indicating the presence of a big-six accounting firm (Titman and Trueman [1986], Beatty [1989], Teoh and Wong [1993]). The variable captures the Big 5 after Price Waterhouse merged with Coopers & Lybrand, and then the Big 4 after Arthur Andersen's demise.

Market conditions

Average Prior Three-month Underpricing = Average percent of underpricing of IPOs in the prior three months (Ljungqvist [2007]).

Three-week Prior Market Return = Percent return on the market (value-weighed NYSE/AMEX/NASDAQ) three weeks (15 trading days) prior to the IPO (Loughran and Ritter [2002]).

Table 5. Summary statistics for dependent and independent variables used in regressions of IPO mispricing.

Variable	Mean	Deviation	Median	Minimum	Maximum
First-day Return (Underpricing)	0.313	0.611	0.121	−0.331	6.975
Offering Characteristics					
IPO Proceeds ($ million)	134.5	323.2	71.7	5.8	8,903.6
Pre-offer Price Change (%)	2.81	27.92	0.00	−98.44	220
Underwriter Spread (%)	6.91	0.60	7.00	2.73	10.00
Refinance Debt (%)	19.30	39.47	0.00	0.00	1.00
Number of Amendments	4.07	1.79	4.00	0	13
Secondary Shares Offered (%)	7.64	15.48	0.00	0	88.61
Dual Share Class (%)	8.63	28.08	0	0	1
High Tech Dummy (%)	38.91	48.76	0	0	1
Internet IPO Dummy (%)	18.10	38.51	0.00	0	1
Third-Party Certification					
Underwriter Prestige	7.86	1.69	8.40	0	9.001
Venture Capital Backing Dummy	0.49	0.50	0	0	1
Big-Six Auditor Dummy	0.81	0.39	1	0	1
Market Conditions					
Average Prior 3-month Underpricing (%)	30.69	26.91	16.77	−19.90	105.77
Three-week Prior Market Return (%)	0.84	3.54	1.20	−13.23	14.34
Recent IPO Volume	115.9	64.1	111.0	3.0	248.0

Sample of 2,247 IPOs from 1996 to 2008. We retrieve the registration documents from EDGAR; the original list of IPOs, IPO proceeds, offer price, pre-offer price change, number of amendments, secondary shares offered, use of proceeds, underwriter name, underwriter spread, VC status, and auditor name from SDC; the first-day closing price and three-week prior market return from CRSP; and the book value of equity from Compustat. We also use the internet dummy status, underwriter rank, firm founding date, dual-class status, number of IPOs in the prior three months, and average prior three-month underpricing from Jay Ritter's webpage. Three-week prior market return is the CRSP value-weighted NYSE/AMEX/NASDAQ return.

Recent IPO volume = Sum of gross IPOs in the prior three months (Welch [1992]).

Additionally, we include year dummy variables and industry dummy variables as in the research by Benveniste, Ljungqvist, Wilhelm, and Yu [2003].

Table 5 reports summary statistics for our dependent and independent variables. The average IPO experienced 31.3% underpricing, with one firm's (VA Linux Systems, December 1999) stock increasing by nearly 700%. The 15 control variables measuring offering characteristics indicate that our sample firms float an average $135 million offering, adjust the pre-offer price range up 2.8%, have a 7% underwriter spread and four amendments prior to the issue, and sell 7.6% of the issue as secondary shares. Only 19% use a portion of the proceeds to refinance debt, 9% have dual share classes, 39% are high tech firms, and 18% Internet firms. The three third-party certification variables indicate an average underwriter prestige score of 7.86 (on a range of 0 to 9.001), 49% with venture capital backing, and 81% audited by a big-six accounting firm. The final set of control variables (i.e., market conditions) indicate that the prior three-month average underpricing was nearly 31%, the three-week prior market return average was 84 basis points, with 116 IPOs in the prior three months.

First-day return results

Base-case results

The Table 3 Pearson correlation coefficient between our measure of strategic tone and IPO first-day return is 0.2401 (p-value less than 0.0001). Soft information, specifically the *Strategic Tone* of the registration statement, is univariately correlated with the IPO pricing anomaly. Furthermore, the (unreported) correlation coefficient between our negative strategy score, *Negative Score*, and *Return* is −0.1293 (p-value less than 0.0001), and the correlation coefficient between our measure of positive words, *Positive Score*, and *Return* is 0.1383 (p-value less than 0.0001). Thus, from a pair-wise correlation point of view, the frequency of positive and negative strategy words helps explain the IPO underpricing. These three significant correlations could potentially be driven by a known determinant of underpricing. For example, positive strategic language may be more prevalent when IPOs are abundant and recent first-day returns are large. Consequently, we next test the correlation between strategic tone and first-day return in a multivariate setting.

Our regression equation for this test is:

$$Return_i = \alpha_0 + \alpha_1 Strategic\ Tone_i + \sum_{j=1}^{15} \beta_j x_{j,i} + \sum_{t=1}^{12} \gamma_t Y_{t,i} + \sum_{k=1}^{6} \delta_k SIC_{k,i} + e_i, \tag{4}$$

where x_j are the 15 control variables starting with *IPO proceeds* and ending with *Recent IPO Volume;* Y_t are year dummies with 1999 omitted; and SIC_k are industry dummies with the services industry omitted. Our main focus is on the null hypothesis: H_0: $\alpha_1 = 0$. If the IPO underwriters ignore a document's value-relevant information embedded within its strategic words, or if investors are temporarily swayed by a document's strategic buzz words, then strategic tone matters and $\alpha_1 > 0$.

In Model 1 of Table 6, we first keep *Strategic Tone* separated into *Negative Score* and *Positive Score* and include an interaction term, *Negative Score* * *Positive Score*. The insignificance of the interaction term (coefficient = 0.006, *p*-value 0.4714) indicates that an additive model is appropriate. For example, *Positive Score*'s effect on returns is independent of the *Negative Score* level. Furthermore, the effect of negative and positive strategic words is somewhat symmetric. For the negative word score, a registration document that has a frequency of negative strategic words one standard deviation higher has a 3.1% *lower* first-day return (*p*-value = 0.0097), whereas a frequency of positive strategic words one standard deviation higher is correlated with 2.5% *higher* first-day return (*p*-value = 0.0256). An *F*-test is unable to reject symmetry; specifically, a null hypothesis stating that the coefficient on *Positive Score* equals the absolute value of the coefficient of *Negative Score* cannot be rejected (*p*-value = 0.4089). Thus, the two-dimensional space of positive and negative strategic words statistically collapses down to the single dimension of overall strategic tone, and we can combine the positive and negative scores using the simple subtraction in Equation 3 without imposing an erroneous restriction on the two slope coefficients.

Model 2 in Table 6 reports the estimation of Equation 4. The regression results document that the null hypothesis stating that soft information is irrelevant (H_0: $\alpha_1 = 0$) is rejected with a *p*-value of 0.0066. The single measure of strategic tone is statistically significant. A registration document with a one standard deviation higher strategic tone has 2.7% *higher* first-day return, on average. From Table 5, the median IPO's price increases 12.1% on the first day. Thus, an IPO's prospectus with a strategic tone of 1 would have an average first-day return of 12.1 + 2.7 = 14.8%. Thus, the IPO market is inefficient with respect to publicly available words in the prospectus.

Model 3 of Table 6 reports the results of our survey-based weighting system. Relative to the unweighted results in Model 2, the coefficient is slightly smaller (i.e., 0.021)

Table 6. Coefficient estimates of IPO mispricing on measures of the registration documents' strategic tone.

	Model 1		Model 2		Model 3		Model 4		Model 5	
Variable	Estimate	*p*-value	Estimate	*p*-value	Estimate	*p*-value	Estimate	*p*-value	Estimate	*p*-value
Negative Score	−0.031	0.0097								
Positive Score	0.025	0.0256								
Negative Score * *Positive Score*	0.006	0.4714								
Strategic Tone			0.027	0.0066			0.031	0.0025		
Weighted Tone					0.021	0.0277				
L&M Fin-Negative							0.026	0.0090	0.025	0.0126
L&M Fin-Positive							−0.009	0.3822	−0.001	0.9154
IPO Proceeds (log)	−0.060	0.0016	−0.061	0.0015	−0.063	0.0009	−0.055	0.0043	−0.062	0.0012
Pre-Offer Price Change	0.012	<.0001	0.012	<.0001	0.012	<.0001	0.012	<.0001	0.012	<.0001
Underwriter Spread	0.016	0.4831	0.015	0.4948	0.015	0.5038	0.018	0.4241	0.015	0.5008
Refinance Debt	−0.027	0.3065	−0.027	0.3058	−0.028	0.2875	−0.028	0.2920	−0.033	0.2092
Number of Amendments	0.014	0.0163	0.014	0.0158	0.014	0.0119	0.013	0.0227	0.014	0.0152
Secondary Shares Offered	−0.001	0.0443	−0.001	0.0464	−0.001	0.0579	−0.001	0.0511	−0.001	0.0875
Dual Share Class	0.009	0.7997	0.008	0.8130	0.010	0.7667	0.012	0.7422	0.016	0.6437
High Tech Dummy	0.005	0.8142	0.005	0.8270	0.010	0.6688	0.000	0.9968	0.013	0.5818
Internet IPO Dummy	0.172	<.0001	0.172	<.0001	0.175	<.0001	0.165	<.0001	0.169	<.0001
Underwriter Prestige	0.021	0.0022	0.021	0.0023	0.021	0.0022	0.021	0.0026	0.022	0.0014
Venture Capital Backing Dummy	0.042	0.0584	0.041	0.0588	0.046	0.0365	0.036	0.1025	0.044	0.0457
Big Six Auditor Dummy	−0.045	0.0692	−0.045	0.0683	−0.046	0.0636	−0.046	0.0625	−0.048	0.0524
Average Prior 3-Month Underpricing	0.002	0.0071	0.002	0.0074	0.002	0.0078	0.002	0.0055	0.002	0.0071
Three-week Prior Market Return	0.005	0.0853	0.005	0.0853	0.005	0.0791	0.005	0.0930	0.005	0.0785
Recent IPO Volume	0.000	0.3729	0.000	0.3860	0.000	0.3684	0.000	0.4099	0.000	0.4003
Intercept	1.030	0.0222	1.048	0.0198	1.094	0.0149	0.931	0.0393	1.071	0.0174
Years	Yes		Yes		Yes		Yes		Yes	
Industries	Yes		Yes		Yes		Yes		Yes	
Adjusted R^2	48.46%	<.0001	48.46%	<.0001	48.40%	<.0001	48.60%	<.0001	48.41%	<.0001

Sample of 2,247 IPOs from 1996–2008. The coefficients are from our Equation 4 regression:

$$Return_i = \alpha_0 + \alpha_1 Strategic\ Tone_i + \sum_{j=1}^{15} \beta_j x_{j,i} + \sum_{t=1}^{12} \gamma_t Y_{t,i} + \sum_{k=1}^{6} \delta_k SIC_{k,i} + e_i$$

Where *Return* is the IPO's first-day return, x_j are the 15 control variables starting with *IPO proceeds* and ending with *Recent IPO Volume.* Y_t are year dummies with 1999 omitted; and SIC_k are industry dummies with the services industry omitted. In Model 1, we separate *Strategic Tone* into its two components—the relative frequency of negative words, *Negative Score*, and the relative frequency of positive words, *Positive Score*—and include an interaction term. In Model 3, words are weighted based on their rating from the survey. Models 4 and 5 include the two L&M scores based on Loughran and McDonald's [2011] lists of negative and positive finance words.

and less significant (p-value = 0.0277). Given the appeal of allowing certain words to be more negative or positive in tone than others, the weighted results suggest that the correlation between strategic tone and first-day returns is not quite as strong as our traditional unweighted results indicate. Nevertheless, Model 3 confirms a statistically significant correlation between the weighted tone of a registration document and the first-day IPO return.

Before reporting on the robustness of our finding, we comment on the overall significance of our model and the control variables. Collectively, the independent variables explain just under half of the variation in first-day returns (adjusted-$R^2 = 0.4846$ in Model 2). Our set of control variables is generally consistent with the extant literature. *IPO Proceeds, Secondary Shares Offered*, and *Big Six Auditor* are all negatively correlated with return, consistent with Habib and Ljungqvist [1998], Leland and Pyle [1977], and Titman and Trueman [1986], respectively. *Pre-Offer Price Change, Number of Amendments, Internet IPO Dummy, Underwriter Prestige, Average Prior Three-Month Underpricing*, and *Recent IPO Volume* are all positively correlated with return, consistent with Hanley [1993], Ang and Brau [2003], Loughran and Ritter [2004], Beatty and Welch [1996], Brav and Gompers [1997], Ljungqvist [2007], and Welch [1992], respectively.

Several control variables in Model 2 of Table 6 are insignificant. These insignificant control variables substantiate the practical importance of our finding regarding the new strategic soft variable. The finance literature once considered all of these control variables important determinants of IPO mispricing. That is, *Underwriter Spread, Refinance Debt, Dual Share Class, High Tech Dummy*, and *Venture Capital Backing Dummy* were (at least) once considered important. We find that our new strategic tone variable is statistically significant whereas these previously "important" variables are not.

For comparison purposes, Models 4 and 5 of Table 6 report the results of using Loughran and McDonald's [2011] list of finance words with and without our score. Model 4 indicates that the addition of the Loughran and McDonald [2011] variables has no significant effect on our measure of strategic tone; in fact, our finding is slightly strengthened with their addition. Likewise, comparing Model 5 with Model 4 indicates that strategic tone has little influence on the finance variable coefficients. As in Ferris et al. [2013] and Loughran and McDonald [2013], we find that negative finance words increase the size of the first-day return. A document with one standard deviation higher score of negative finance word frequency is associated with 2.5% higher first-day return (p-value = 0.0126).[5]

Potentially our strategy results could be another manifestation the Loughran and McDonald's [2013] finding about uncertainty if somehow our measure of strategic tone and their measure of negative finance were both actually measuring uncertainty. We believe this is not the case. First, we exclude Loughran and McDonald words from our library making it difficult for the separate word lists to capture the same content. Second, the correlation coefficient between our strategic tone and their negative finance variables is only 0.06. Third, our strategy finding is symmetric—positive words increase the first day return about the same amount as negative words decrease it. Such is not the case with Loughran and McDonald's proxies for uncertainty. They find negative, but not positive finance words, and weak modal, but not strong modal, correlate with the first day return.

Robustness

In Table 6 we made choices regarding measurement, dates, and dependent variables. In Table 7 we examine the robustness of the strategic tone coefficient to these choices as well as document the importance of how tone is measured. The first row of Table 7 repeats the base case, Model 2, α_1 coefficient for comparison purposes.

When we divide our sample into two subperiods, 1996–2000 and 2001–2008, the coefficient on *Strategic Tone* is significant in both subsamples. The later sample has both a lower coefficient and a higher p-value, but unreported Chow tests support coefficient stability. For example, the early years' coefficient on *Strategic Tone* is 3.8% and the average underpricing over that period is 39.4% for a 9.5% (3.8/39.4) relative explanatory power. The same ratio in the later years is 1.5/11.9 for a relative explanatory power of 12.3%. That is, the absolute effect of tone on underpricing is bigger in the earlier sample; however, during this period, there was much more

Table 7. Robustness tests.

	Est Coef	p-value
Base Case	0.027	0.0066
Early Years (1996–2000)	0.038	0.0086
Late Years (2001–2008)	0.015	0.0173
Bubble Years (1998–2000)	0.070	0.0069
Non-Bubble Years (1996–1997, 2001–2008)	0.021	<.0001
Base Case, without insignificant controls	0.028	0.0042
Base Case, less ambiguous words	0.028	0.0170
Base Case, no negation	0.021	0.0187
Base Case, no symbols	0.027	0.0199
Base Case without DICTION words	0.027	0.0066

Sample of 2,247 IPOs from 1996–2008. The coefficients are from our Equation 4 regression:

$$Return_i = \alpha_0 + \alpha_1 Strategic\ Tone_i + \sum_{j=1}^{15}\beta_j x_{j,i} + \sum_{t=1}^{12}\gamma_t Y_{t,i} + \sum_{k=1}^{6}\delta_k SIC_{k,i} + e_i$$

Where *Return* is the IPO's first-day return, x_j are the 15 control variables; Y_t are year dummies with 1999 omitted; and SIC_k are industry dummies with the services industry omitted. The estimated coefficient and p-values are reported for *Strategic Tone* under various robustness scenarios.

underpricing to explain. Scaling is of particular importance during the bubble years where we see a 7% estimated coefficient; however, the average underpricing during the bubble years is 58.4% for a 12.0% (7/58.4) relative explanatory power ratio. During the nonbubble years, the explanatory power ratio increased to 16.3% (2.1/12.8), indicating that the frothy markets of the bubble years are not driving our results.

The list of 15 control variables in Table 6 consists of the *IPO Return* determinants found by previous researchers. Excluding the five insignificant (p-values > 0.10) independent variables increases the α_1 coefficient from 0.027 to 0.028. Thus, excluding these variables does not affect our rejection of our null hypothesis 1—we still find that *Strategic Tone* is positively correlated with the first-day's stock return.[6] Likewise, expanding the list of control variables to include age and sales (Ritter [1984]) and overhang (Bradley and Jordan [2002]) does not affect our results.

In our base-case results, we include every word with a survey rank above 3.0 on the positive list and 3.0 or below on the negative list although there was occasional disagreement among survey participants on words ranked close to 3. For example, "exception" was ranked 2.91 and "decentralize" was ranked 3.09. Excluding words between 2.9 and 3.1 (potentially ambiguous words) increases the coefficient but slightly weakens the significance; for example, the p-value for *Strategic Tone* increases slightly to 0.0170.

When measuring strategic tone of a document, we also take care to expand the list of negation words and count symbols the same as words (© is counted the same as the word "copyright"). Ignoring negation lowers the α_1 coefficient 22%, from 0.027 to 0.021. The effect of not counting symbols is much less dramatic. Nevertheless, even the unrefined measures of tone that ignore negation and symbols are statistically significant.

When creating our list of strategy words, we err on the side of inclusion; some of our "strategy" words also appear as positive or negative words in DICTION's libraries of "general" words. However, excluding words found on DICTION's list for praise, satisfaction, inspiration, blame, hardship, and denial has little material effect on the results of Table 6.

Long-run results

If publically available data (i.e., words in the IPO registration document) predict first-day returns, then either the underwriters or the investors are inefficient with respect to this soft information. In one scenario, the underwriters' IPO price could efficiently incorporate soft information into the offer price and, as originally claimed by Ritter [1991], investors could be "overoptimistic about the earnings potential" after seeing positive-sounding strategy words in the registration statement. We call this explanation "investor overreaction" because they price companies with positive-sounding strategies too high and price companies with negative-sounding strategies relatively low. In the alternative case, investors could discover or better utilize overlooked soft information related to the firm's strategic position, or in Hoberg's [2007] words, the underwriters could be "low-balling" the offer price of IPOs with strong strategies. We call underwriter mispricing "underreaction," when investment bankers price strong strategies relatively low and price weak strategies too highly. We test whether the part of first-day returns related to publicly available soft information is due to investor overreaction or investment banker underreaction.

We distinguish between investor overreaction and investment banker underreaction by looking at the long-run performance of the IPOs, conditional on the strategic tone of the registration statement. We avoid the long-run abnormal-return debate; we do not test whether our IPO returns are abnormally high or low but instead focus on whether IPOs with positive strategic tone underperform those with negative strategic tone (e.g., Brav, Geczy and Gompers [2000]). Our implicit assumption is that any error in initial pricing will be corrected after several years of company performance and investor scrutiny. If investors see positive strategy words and overreact, then a portfolio of IPOs with high levels of strategic tone in their registration statement should, in the long run, underperform a portfolio of IPOs with low levels of strategic tone. If underwriters underreact to soft information, then a portfolio of IPOs with negative strategic tones in their registration statements should have typical long-run returns after a high first day. We test whether one group of IPOs (top tercile of strategic tone) does better than another group of IPOs (bottom one-third of strategic tone) over time.

To control for variations in the timing and the riskiness of the positive and negative tone IPOs, we match each IPO to a seasoned benchmark firm on the day of the IPO. We combine the methodologies of Ritter [1991], who matches on industry, and Lyon, Barber, and Tsai [1999], who match on size and market-to-book equity ratios, to find these benchmark firms. Any so-called "treachery" in the benchmarks (see Lyon et al. [1999], p. 198) should be the same between our IPOs with a high strategic tone and those with a low strategic tone.

When assigning a match, we choose the firm that has the closest market-to-book ratio among firms that (1) have not issued equity within five years, (2) have the same 4-digit SIC code, and (3) are within plus or minus

30% market capitalization of the IPO firm as in Barber and Lyon [1997]. If needed, we consider 3-, 2-, and 1-digit SIC code matches. If a matching firm de-lists before the end of the estimation period, we replace it with the next closest market-to-book matching firm at the de-list date. If an IPO de-lists, the abnormal return is equal to the buy-and-hold return through the time of de-listing, minus the benchmark buy-and-hold return.

Table 8 shows the results of our long-run performance tests for 1-, 2-, and 3-year return horizons. Our IPO sample ends in 2008 so that we can follow each firm for three years (return data runs through December 2011). Panel A reports the mean and median abnormal returns for the 766 firms in the top one-third of *Strategic Tone*, whereas Panel B reports on the 766 firms in the bottom one-third of *Strategic Tone*. Panel C reports differences between the top- and bottom-tone portfolio abnormal returns, along with p-values testing the null hypothesis that states that the difference in abnormal returns is zero. Our focus is not on Panels A and B; we are agnostic about whether or not IPOs underperform the benchmark. Rather, we are only concerned with the relative performances of our strategy-tone portfolios in Panel C.

Table 8. Long-run returns categorized by IPO documents' strategic tone.

Panel A. Top 1/3 Tone				
Variable	Mean	Median		
1-Year Abn Rtn	−18.4%	−40.4%		
2-Year Abn Rtn	−33.1%	−58.6%		
3-Year Abn Rtn	−41.1%	−61.5%		
Panel B. Bottom 1/3 Tone				
Variable	Mean	Median		
1-Year Abn Rtn	−10.6%	−17.7%		
2-Year Abn Rtn	−23.6%	−40.3%		
3-Year Abn Rtn	−32.0%	−51.7%		
Panel C. Difference Tests (Top-Bottom)				
Variable	Mean	Median	Mean p-value	Median p-value
1-Year Abn Rtn	−7.9%	−22.7%	0.0158	<.0001
2-Year Abn Rtn	−9.5%	−18.3%	0.0260	0.0029
3-Year Abn Rtn	−9.1%	−9.7%	0.0570	0.0658

Sample of 2,247 IPOs from 1996–2008. Abnormal returns are computed using a modified pair-wise benchmark characteristic approach based on Ritter [1991] and Lyon et al. [1999]. When assigning a match, we choose the firm that has the closest market-to-book ratio among firms that (1) have not issued equity within five years, (2) have the same 4-digit SIC code, and (3) are within plus or minus 30% market capitalization of the IPO firm. If needed, we consider 3-, 2-, and 1-digit SIC codes. If a matching firm de-lists before the end of the estimation period, we replace it with the next closest market-to-book matching firm at the de-list date. If an IPO de-lists, the abnormal return is equal to the buy-and-hold return at the time of listing, minus the benchmark buy-and-hold return. Abnormal return data runs through December 2011. We use the following equation:

$AR_T^i = \prod_{t=0}^{T}(1 + r_t^i) - \prod_{t=0}^{T}(1 + r_t^b)$,

where

AR_T^i is the abnormal buy-and-hold return for firm i for months 0 to 12 (or 24 or 36) after going public,

r_t^i is the raw return for firm i in month t after going public (excluding the first day), and

r_t^b is the pair-matched benchmark firm return in month t as described above.

In all six Panel C comparisons (three return horizons times two return measures, mean and median), the portfolio comprising one-third of the sample firms with the lowest tone outperforms the top one-third tone portfolio (higher abnormal returns), indicating that the investors, not the underwriters, initially mispriced the soft information. For example, the *mean* 3-year abnormal top one-third tone portfolio abnormal return is −41.1%, 9.1% lower than the bottom one-third tone portfolio's −32.0% abnormal return. This 9.1% difference in means rejects the null hypothesis of equal performance (p-value = 0.0570). Likewise, the bottom one-third tone *median* abnormal returns is significantly larger than the top one-third tone, nonparametrically rejecting the null in favor of investors overreacting to the soft information in the IPO registration documents and this inefficiency being corrected over time.

We subject the differences in long-run abnormal returns tests (Table 8, Panel C) to the robustness experiments reported in Table 7. Specifically, we look at subsamples and also classify top and bottom terciles after excluding ambiguous words, excluding DICTION words, and ignoring negation and symbols. As with the first-day return results, the long-run tests show that the evidence that low-tone portfolios outperform is generally robust with weaker, but significant, results when negation is ignored and when the bubble years are dropped from the sample. Whereas the first-day return results are weakest (lowest coefficient) in the later years (IPOs between 2001 and 2008), the long-run results are weakest (smallest difference in abnormal returns) in the early years (IPOs between 1996 and 2000). In fact, the only long-run test results that are not robust are for the early years using 2- and 3-year abnormal return horizons.

Parallel with Bradley and Jordan [2002], who document that publicly available *quantitative* information (e.g., file-range amendments) helps explain a stock's first day's return, we document that widely distributed *qualitative* information (e.g., strategic words) does the same. Bradley and Jordan ([2002], p. 596) assume that their evidence "strongly indicates that IPO offer prices do not fully adjust to public information." We compare subsequent long-run portfolio returns and find evidence that offer prices appear to be efficient with respect to soft strategic information, but the market's closing price is inefficient.

Conclusions

Recent papers show that the soft information (information provided by words) in the IPO registration document are correlated with IPO underpricing. Arnold, Fishe, and North [2010] find that a document's ambiguity leads to underpricing. Hanley and Hoberg [2012] find that the lack of so-called "informativeness" or uniqueness in the document leads to underpricing. Ferris, Hao, and Liao [2013] and Loughran and McDonald [2013] find that conservatism and uncertainty, respectively, lead to underpricing. We use content analysis to quantify a document's strategic tone; positive strategic words, such as "innovative" or "sustainable," increase strategic tone, and negative words, such as "rivals" or "obsolete," decrease strategic tone. We then show that strategic tone is correlated with IPO underpricing (the percent difference between the offer price and the price at the end of the first trading day). We make the following contributions:

First, we document that an IPO registration's publicly available strategic tone is positively correlated with the first-day return. IPOs with a higher frequency of positive strategic words and/or a lower frequency of negative strategic words have, on average, a statistically significant higher first-day return. A registration document with a strategic tone that is one standard deviation higher would have a first-day return of 14.8%, 2.7% higher than the median of 12.1%. The statistical significance of our strategic tone variable is greater than many previously documented determinants of mispricing including underwriter spread, secondary shares offered, dual share class indicator, high tech dummy, and underwriter prestige.

The correlation between an IPO's soft strategic information and return could be driven by the underwriter's inefficient use of soft information leading to an IPO offer price that is "too low" with respect to the publicly available strategic content of the registration statement. On the other hand, the correlation could be driven by investors' inefficient use of soft information leading to an end-of-day market price that is "too high."

Our second contribution documents relative long-run performance consistent with the latter explanation—investors appear to initially misprice the soft information, thus adding to the inefficiency of the IPO process. Specifically, a portfolio of IPOs in the top tercile of strategic tone underperforms a portfolio of IPOs in the bottom one-third of strategic tone in the long run. This shows that any tone-related mispricing on the first day is corrected sometime in the following several years. Together, these two contributions imply that investors decrease the efficiency of the first-day price by misusing the document's strategic tone. Thus, the first-day return can be predicted by publicly available soft information, but it is the closing price, not the offer price, that is inefficient with respect to strategy-related words.

Third, we create new content-analysis libraries for positive and negative strategic words and introduce a survey approach to creating libraries. Our survey-based approach to word libraries reduces researcher bias, ranks words in a specific context by business professionals, and generates a continuous rank rather than simple lists of positive and negative word lists. Along the way, we also explore possible improvements to content-analysis methodology and understanding by testing for dimension reduction (combining libraries), developing a new word-weighting scheme, and expanding the list of negation words as well as considering symbols (such as trademark (^{TM})).

Acknowledgments

We appreciate the comments of Michael Hendron, Kathleen Hanley, Roderick Hart, Gerard Hoberg, Tim Loughran, Bill McDonald, and Jay Ritter as well as seminar participants at the University of Washington, BYU, and Rochester Institute of Technology Finance Departments. We thank Eric DeRosia, Glen Christensen, and Qualtrics for help on the survey. McQueen received financial support from the William Edwards Professorship.

Funding

Brau recognizes funding from a Marriott School research grant. We acknowledge the BYU Silver Fund that paid for databases and research support.

Notes

1. In addition to the finance literature highlighted in Table 1, see Li [2010] for a review of textual analysis in the accounting literature.
2. Individual professionals rank words on a discrete scale from 1 to 5; however, the average across all individuals yields continuous scores.
3. The question of whether the two-dimensional space of optimism and pessimism collapses down to one dimension is relevant. Davis et al. [2008], for example, consider an earnings press release's degree of optimism and pessimism as two separate dimensions, whereas Rogers et al. [2009] subtract the three pessimism scores from the three optimism scores for earnings announcements to get one measure of "Net-mism."
4. For three reasons, we do not use the weighting systems from the information retrieval literature—retrieving documents from repositories. First, the goal of retrieval is to find the document(s) most related to a list of key words rather than comparing and contrasting the semantic content of all documents. Second, as documented by Salton and Buckley [1988] and Pincombe [2004], researchers can search across a number of retrieval weighting schemes. In contrast, our weights come from disinterested but expert survey participants. Third, while appropriate for retrieving, common retrieval weighting systems result in

problematic weights when applied to measuring tone. For example, using the popular inverse document frequency function, a word such as "centralize," which survey participants consider only slightly positive receives a much bigger weight than a word such as "leader," which is almost universally associated with a very good strategy by our survey participants. Similarly, global weighting actually results in a word such as "fast" getting a greater weight than "faster," which may bias results.

5. Ferris et al.'s [2013] interpretation is that negative finance words are a proxy for conservatism, which leads to underpricing. Loughran and McDonald's [2013] interpretation is that negative finance words are a proxy for uncertainty which leads to underpricing.
6. Excluding the insignificant control variables keeps Underwriter Prestige and excludes Underwriter Spread avoiding possible multicollinearity. James [1992] and Fang [2005] find that these two variables are negatively correlated.

References

Ang, J., and J. Brau. "Concealing and Confounding Adverse Signals: Insider Wealth-maximizing Behavior in the IPO Process." *Journal of Financial Economics*, 67, (2003), pp. 149–172.

Antweiler, W., and M. Z. Frank. "Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards." *Journal of Finance*, 59, (2004), pp. 1259–1293.

Arnold, T., R. Fishe, and D. North. "The Effects of Ambiguous Information on Initial and Subsequent IPO Returns." *Financial Management*, 39, (2010), pp. 1497–1519.

Baginski, S., E. Demers, C. Wang, and J. Yu. "Understanding the Role of Language in Management Forecast Press Releases." INSEAD, Working Paper (2011).

Barber, B., and J. Lyon. "Detecting Long-run Abnormal Stock Returns: The Empirical Power and Specification of Test Statistics." *Journal of Financial Economics*, 42, (1997), pp. 341–372.

Baron, D. "A Model of the Demand for Investment Banking Advising and Distribution Services for New Issues." *Journal of Finance*, 37, (1982), pp. 955–976.

Barry, C., C. Muscarella, J. W. Peavy III, and M. Vetsuypens. "The Role of Venture Capital in the Creation of Public Companies: Evidence from the Going Public Process." *Journal of Financial Economics*, 27, (1990), pp. 447–471.

Beatty, R. P. "Auditor Reputation and the Pricing of Initial Public Offerings." *Accounting Review*, 64, (1989), pp. 693–709.

Beatty, R. P., and J. R. Ritter. "Investment Banking, Reputation, and the Underpricing of Initial Public Offerings." *Journal of Financial Economics*, 15, (1986), pp. 213–232.

Beatty, R. P., and I. Welch. "Issuer Expenses and Legal Liability in Initial Public Offerings." *Journal of Law and Economics*, 39, (1996), pp. 545–602.

Benveniste, L. M., A. P. Ljungqvist, W. J. Wilhelm Jr., and X. Yu. 2003. "Evidence of Information Spillovers in the Production of Investment Banking Services." *Journal of Finance*, 58, (2003), pp. 577–608.

Bettman, J. R., and B. A. Weitz. "Attributions in the Board Room: Causal Reasoning in Corporate Annual Reports." *Administrative Science Quarterly*, 28, (1983), pp. 165–183.

Booth, J. R., and R. L. Smith II. "Capital Raising Underwriting and the Certification Hypothesis." *Journal of Financial Economics*, 15, (1986), pp. 261–281.

Bradley, D. J., and B. D. Jordan. "Partial Adjustment to Public Information and IPO Underpricing." *Journal of Financial and Quantitative Analysis*, 37, (2002), pp. 595–616.

Brav, A., and P. A. Gompers. "Myth or Reality? The Long-run Underperformance of Initial Public Offerings: Evidence From Venture and Nonventure Capital-backed Companies." *Journal of Finance*, 52, (1997), pp. 1791–1821.

Brav, A., C. Geczy, and P. A. Gompers. "Is the Abnormal Return Following Equity Issuances Anomalous?" *Journal of Financial Economics*, 56, (2000), pp. 209–249.

Brockman, P., and J. Cicon. "The Information Content of Management Earnings Forecasts: An Analysis of Hard Versus Soft Information." *Journal of Financial Research*, 36, (2013), pp. 147–174.

Carter, R. B., F. H. Dark, and A. K. Singh. "Underwriter Reputation, Initial Returns, and the Long-run Performance of IPO Stocks." *Journal of Finance*, 53, (1998), pp. 285–311.

Carter, R. B., and S. Manaster. "Initial Public Offerings and Underwriter Reputation." *Journal of Finance*, 45, (1990), pp. 1045–1067.

Chen, H., and J. R. Ritter. "The Seven Percent Solution." *Journal of Finance*, 55, (2000), pp. 1105–1131.

Cicon, J., J. Clarke, S. P. Ferris, and N. Jayaraman. "Managerial Expectations and the Performance of Acquiring Firms: The Contribution of Soft Data." University of Missouri, Working Paper (2010).

Das, S. R., and M. Y. Chen. "Yahoo! For Amazon: Sentiment Extraction from Small Talk on the Web." *Management Science*, 53, (2007), pp. 1375–1388.

Davis, A. K., J. Piger, and L. M. Sedor. "Beyond the Numbers: Managers' Use of Optimistic and Pessimistic Tone in Earnings Press Releases." University of Oregon, Working Paper (2008).

Demers, E., and C. Vega. "Soft Information in Earnings Announcements: News or Noise?" INSEAD, Working Paper (2008).

Engelberg, J. "Costly Information Processing: Evidence from Earnings Announcements." Northwestern University, Working Paper (2008).

Fang, L. "Investment Bank Reputation and the Price of Quality of Underwriting Services." *Journal of Finance*, 60, (2005), pp. 2729–2761.

Ferris, S., Q. Hao, and M. Liao. "The Effect of Issuer Conservatism on IPO Pricing and Performance." *Review of Finance*, 17, (2013), pp. 933–1027.

Field, L., and G. Hanka. "The Expiration of IPO Share Lockups." *Journal of Finance*, 56, (2001), pp. 471–500.

Gurun, U., and A. Butler. "Don't Believe the Hype: Local Media Slant, Local Advertising, and Firm Value." *Journal of Finance*, 62, (2012), pp. 561–598.

Habib, M. A., and A. P. Ljungqvist. "Underpricing and IPO Proceeds: A Note." *Economics Letters*, 61, (1998), pp. 381–383.

Hanley, K. W. "The Underpricing of Initial Public Offerings and the Partial Adjustment Phenomenon." *Journal of Financial Economics*, 34, (1993), pp. 231–250.

Hanley, K. W., and G. Hoberg. "The Information Content of IPO Prospectuses." *Review of Financial Studies*, 23, (2010), pp. 2821–2864.

Hanley, K. W., and G. Hoberg. "Are Strategic Disclosure and Underpricing Decisions Influenced by Litigation Risk?" *Journal of Financial Economics*, 103, (2012), pp. 235–254.

Hart, R. P. "Systematic Analysis of Political Discourse: The Development of Diction." In K. Sanders et al., eds., *Political Communication Yearbook* (pp. 97–134.) Carbondale: Southern Illinois University Press, 1984.

Hart, R. P. *DICTION 5.0: The Text-analysis Program*. Thousand Oaks, CA: Scolari/Sage, 2000.

Hart, R. P. "Redeveloping DICTION: Theoretical Consideration." In M. West, ed., *Theory, Method, and Practice of Computer Content Analysis*. New York: Ablex Publishing, 2001.

Hitt, M. A., R. D. Ireland, and R. E. Hoskisson. *Strategic Management: Competitiveness and Globalization*. New York: South-Western, 2005.

Hoberg, G. "The Underwriter Persistence Phenomenon." *Journal of Finance*, 62, (2007), pp. 1169–1206.

Ibbotson, R. G. "Price Performance of Common Stock New Issues." *Journal of Financial Economics*, 2, (1975), pp. 235–272.

James, C. "Relationship-specific Assets and the Pricing of Underwriter Services." *Journal of Finance*, 47, (1992), pp. 1865–1885.

Jegadeesh, N., and D. Wu. "Word Power: A New Approach for Content Analysis." Emory University, Working Paper (2011).

Kahle, K. M., and R. A. Walkling. "The Impact of Industry Classifications on Financial Research." *Journal of Financial and Quantitative Analysis*, 31, (1996), pp. 309–336.

Kothari, S. P., X. Li, and J. E. Short. "The Effect of Disclosures by Management, Analysts, and Financial Press on Cost of Capital, Return Volatility, and Analyst Forecasts: A Study Using Content Analysis." *Accounting Review*, 84, (2009), pp. 1639–1670.

Krippendorf, K. *Content Analysis: An Introduction to Its Methodology*, 2nd edition. Thousand Oaks, CA: Sage, 2004.

Leland, H. E., and D. H. Pyle. "Informational Asymmetries, Financial Structure, and Financial Intermediation." *Journal of Finance*, 32, (1977), pp. 371–387.

Li, F. "Do Stock Market Investors Understand the Risk Sentiment of Corporate Annual Reports?" (April 21, 2006). Retrieved from http://ssrn.com/abstract=898181.

Li, F. "Textual Analysis of Corporate Disclosures: A Survey of the Literature." *Journal of Accounting Literature*, 29, (2010), pp. 143–165.

Ljungqvist, A. "IPO Underpricing." In B. Espen Eckbo, ed., *Handbook of Corporate Finance: Empirical Corporate Finance, Volume 1* (pp. 375–421). North-Holland: Elsevier, 2007.

Logue, D. E. "Premia on Unseasoned Equity Issues: 1965–1969." *Journal of Financial and Quantitative Analysis*, 25, (1973), pp. 133–141.

Loughran, T., and B. McDonald. "When is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks." *Journal of Finance*, 66, (2011), pp. 35–65.

Loughran, T., and B. McDonald. "IPO First-day Returns, Offer Price Revisions, Volatility, and Form S-1 Language." *Journal of Financial Economics*, 109, (2013), pp. 307–326.

Loughran, T., and J. R. Ritter. "Why Don't Issuers Get Upset About Leaving Money on the Table in IPOs?" *Review of Financial Studies*, 15, (2002), pp. 413–443.

Loughran, T., and J. R. Ritter. "Why Has IPO Underpricing Changed Over Time?" *Financial Management*, 33, (2004), pp. 5–37.

Lyon, J. D., B. M. Barber, and C. Tsai. "Improved Methods for Tests of Long-run Abnormal Stock Returns." *Journal of Finance*, 54, (1999), pp. 165–201.

Megginson, W. L., and K. A. Weiss. "Venture Capital Certification in Initial Public Offerings." *Journal of Finance*, 46, (1991), pp. 879–904.

Ober, S., J. J. Zhao, R. Davis, and M. Alexander. "Telling It Like It Is: The Use of Certainty in Public Business Discourse." *Journal of Business Communication*, 36, (1999), pp. 280–300.

Pincombe, B. *Comparison of Human and Latent Semantic Analysis (LSA) Judgments of Pairwise Document Similarities for a News Corpus*. Edinburgh: DSTO Information Sciences Laboratory, 2004.

Porter, M. "An Algorithm for Suffix Stripping." *Program*, 14, (1980), pp. 130–137.

Price, S., J. Doran, D. Peterson, and B. Bliss. "Earnings Conference Calls and Stock Returns: The Incremental Informativeness of Textual Tone." *Journal of Banking and Finance*, 36, (2012), pp. 992–1011.

Ritter, J. R. "Signaling and the Valuation of Unseasoned New Issues: A Comment." *Journal of Finance*, 39, (1984), pp. 1231–1237.

Ritter, J. R. "The Long-run Performance of Initial Public Offerings." *Journal of Finance*, 46, (1991), pp. 3–27.

Rock, K. "Why New Issues are Underpriced." *Journal of Financial Economics*, 15, (1986), pp. 187–212.

Rogers, J. L., A. Van Buskirk, and S. L. C. Zechman. "Disclosure Tone and Shareholder Litigation." University of Chicago, Working Paper (2009).

Salton, G., and C. Buckley. "Term Weighting Approaches in Automatic Test Retrieval." *Information Processing Management*, 25, (1988), pp. 513–523.

Smart, S. B., and C. J. Zutter. "Control as a Motivation for Underpricing: A Comparison of Dual and Single-class IPOs." *Journal of Financial Economics*, 69, (2003), pp. 85–110.

Stoll, H. R., and A. J. Curley. "Small Business and the New Issues Market for Equities." *Journal of Financial and Quantitative Analysis*, 5, (1970), pp. 309–322.

Teoh, S. H., and T. J. Wong. "Perceived Auditor Quality and the Earnings Response Coefficient." *Accounting Review*, 68, (1993), pp. 346–366.

Tetlock, P. C. "Giving Content to Investor Sentiment: The Role of Media in the Stock Market." *Journal of Finance*, 62, (2007), pp. 1139–1168.

Tetlock, P. C., M. Saar-Tsechansky, and S. Macskassy. "More Than Words: Quantifying Language to Measure Firms' Fundamentals." *Journal of Finance*, 63, (2008), pp. 1437–1467.

Titman, S., and B. Trueman. "Information Quality and the Valuation of New Issues." *Journal of Accounting and Economics*, 8, (1986), pp. 159–172.

Welch, I. "Sequential Sales, Learning and Cascades." *Journal of Finance*, 47, (1992), pp. 695–732.

West, M. D. (Ed.). *Theory Method and Practice of Computer Content Analysis*. New York: Ablex Publishing, 2001.

Appendix

We begin with 5,689 strategy-related words from Hitt, Ireland, and Hoskisson's [2005] *Strategic Management* and DICTION. After deleting words that are also on Loughran and McDonald's [2011] lists of legal and financial words, we are left with 2,349 words. Among these remaining words, the following table reports the top-10 and bottom-10 words by average student survey rating, with 1 indicating a very poor business strategy and 5 indicating a very good business strategy.

Top ten words by rating		Bottom ten words by rating	
Word	Rating	Word	Rating
exemplary	4.460	bribing	1.220
leader	4.450	steal	1.279
premium	4.349	scandal	1.293
pioneer	4.341	incapability	1.400
optimize	4.302	miscalculating	1.436
differentiate	4.293	felon	1.439
agile	4.220	embezzle	1.447
proprietary	4.209	launder	1.463
dependably	4.205	harass	1.468
unique	4.200	sabotaging	1.475

In the above table, we list only one word from a word group. For example, "bribing" and "bribability" are both part of the word group "bribe," that receives a 1.22 survey ranking.