Hyperbole or Reality? Investor Response to Extreme Language in Earnings Conference Calls

Khrystyna Bochkay[*] Sudheer Chava[†] Jeffrey Hales[‡]

Abstract

We develop a dictionary of linguistic extremity in earnings conference calls, a setting where managers have considerable latitude in the language they use, to study the role of extreme language in corporate reporting. Controlling for tone (positive vs. negative) of language, we document that when managers use more extreme words in earnings conference calls, trading volume around the call increases and stock prices react more strongly. In addition, both effects are more pronounced for firms with weaker information environments. Linguistic extremity also affects analyst opinions and contains information about a firm's future operating performance. As such, our results provide evidence that markets are influenced not just by *what* managers say, but also *how* they say it, with extreme language playing an important role in communicating reality and not merely reflecting hyperbole.

Keywords: Extreme Language; Market Reactions; Analyst Forecast Revisions; Future Performance; Earnings Conference Calls; Textual Analysis.

[*]Khrystyna Bochkay: School of Business Administration, University of Miami, Coral Gables, FL 33146. Email: *kbochkay@bus.miami.edu*. Phone: (305) 284-2334.

[†]Sudheer Chava: Scheller College of Business, Georgia Institute of Technology, Atlanta, GA 30308. Email: *sudheer.chava@scheller.gatech.edu*. Phone: (404) 894-4371.

[‡]Jeffrey Hales: Scheller College of Business, Georgia Institute of Technology, Atlanta, GA 30308. Email: *jeffrey.hales@scheller.gatech.edu*. Phone: (404) 894-3897.

We would like to thank Pietro Bianchi, Mark Bradshaw, Kimball Chapman, Roman Chychyla, Valentin Dimitrov, Helena Isidro, Alok Kumar, Andy Leone, Carolyn Levine, Tim Loughran, Andrew McMartin, Miguel Minutti-Meza, Dhananjay Nanda, Sundaresh Ramnath, Eric Weisbrod, Mike Willenborg, Sunny Yang, seminar participants at the University of Miami, University of Connecticut, conference participants at the 2016 EAA Annual Congress, the 2017 FARS Midyear Meeting, the 2017 INSEAD Accounting Symposium, the Behavioral Finance Seminar participants at the Georgia Institute of Technology and University of Bern for their helpful comments and suggestions.

"Don't say infinitely when you mean very; otherwise you'll have no word left when you want to talk about something really infinite." C. S. Lewis

1 Introduction

If a manager were to describe a corporate event or operating results using relatively more or less extreme language, would the market notice and, if so, what should investors read into managers' choice of words?[1] In recent years, there has been considerable interest in using linguistic analysis to better understand managers' use of and markets reactions to non-numerical (linguistic) corporate disclosures (see Li (2010b) and Loughran and McDonald (2016) for a review). Unlike numbers, which are subject to formal accounting rules, language brings with it an infinite number of possibilities. Even when thinking about a single concept or thought, the number of ways in which that thought might be expressed is seemingly boundless, and this is no less true in the domain of corporate communication than in interpersonal communication. For example, positive earnings growth could be described as "surpassing" analyst expectations or as "soaring" beyond those expectations. In this paper, we argue that an important attribute of corporate communication is the linguistic extremity of the words management chooses to use.[2]

Our view is informed by numerous studies in psychology that examine how linguistic extremity and vividness influence decision-making and persuasion (e.g., Burgoon et al. (1975), Nisbett and Ross (1980), Aune and Kikuchi (1993)). In the context of financial reporting, Hales et al. (2011) demonstrate that investment positions can influence whether investors attend to differences in linguistic characteristics, like vividness or extremity, when revising their beliefs about future performance. However, these studies examine individual behavior and do so in experimental settings rather than in naturally occurring corporate settings, and cannot speak to how extreme language affects market-level behaviors, such as abnormal

[1]For example, John Legere, CEO of T-Mobile, makes strong remarks about T-Mobile's performance in recent Q4 2017 earnings conference call: "it's been 5 full years since I called BS and declared war in the status quo of the wireless industry... I broke the news that T-Mobile is going to take a stand against the stupid, broken, arrogant industry... We have great financial results ... very strong free cash flow ... and 2018 will be our best year yet... As you can tell, we're incredibly confident in our future."

[2]Following Bowers (1963), we define *linguistic extremity* as "the quality of language which indicates the degree to which the speaker's attitude toward a concept deviates from neutrality."

trading volume and stock returns. In this paper, we extend this research and contribute, more generally, to the larger literature on linguistic analysis, by examining managers' word choices in earnings conference calls. In particular, we examine whether market participants attend to, and find credible, language that is relatively more extreme. We also test whether extreme language is informative about future operating performance.

On one hand, market participants might be largely skeptical of extreme language, particularly when optimistic, because managers have wide latitude in how they describe their past or current performance (absent any material misstatement of fact) and, if things later turn south, managers may claim in court that their prior statements were mere "puffery."[3] In addition, legislated safe harbor provisions provide sweeping protections to managers when describing future performance, as long as those forward-looking statements are accompanied by meaningful cautionary disclaimers (Asay and Hales, 2018). As such, investors might view the nuances in managers' choice of words as cheap talk and pay little attention to such distinctions. Moreover, even if managers are intending to be truthful, language is inherently ambiguous, and there is no guarantee that the meaning intended will be the meaning received. Thus, even if investors attend to the specific word choices that managers make, those word choices might simply provoke more disagreement among investors, generating trade, but doing little to aid price discovery. Still, if extreme language is viewed as credible and meaningful, we should see markets reacting more strongly when it appears in management communications.

Our analysis is, therefore, structured in three related steps. First, we create a dictionary of words and phrases to measure linguistic extremity. Second, we use this dictionary to test whether investors and analysts react differently to extreme vs. more moderate language in earnings conference calls. Third, we examine whether extreme language in earnings calls carries information about future fundamental performance. Finally, we test whether our measures of linguistic extremity and measures of positive vs. negative tone from prior literature capture the same or different attributes of earnings conference calls.[4]

[3] *Puffery* has been described as "statements that are so optimistic, general, broad, or vague" that courts may view them as not actionable (Osovsky (2016), p. 339).

[4] By using market reactions and future performance tests, we assume that markets are efficient and that management cannot influence future operating results. As a result, all our inferences should be interpreted in the light of these assumptions.

We examine linguistic extremity in the context of earnings conference calls for a number of reasons. First, conference calls represent one of the major forms of investor communication that firms use to supplement the information contained in their financial statements and other regulatory filings (Frankel et al. (1999), Kimbrough (2005), Frankel et al. (2010), Matsumoto et al. (2011)). Second, like many regulatory filings, conference calls contain both numerical and textual information. However, in contrast to the formal and often boilerplate language often seen in regulatory filings, conference calls involve spoken, rather than written, language and so tend to be less formal and scripted than what is typically seen in regulatory filings, such as annual and quarterly SEC reports. As such, the range of words used in conference calls is likely wider, making it a good setting to study managerial choice of language and the impact that linguistic extremity can have on investors' interpretation of information. Lastly, because a large sample of conference call transcripts is available, we can use the text from these transcripts to develop a comprehensive dictionary of linguistic extremity in the context of corporate reporting. In doing so, our primary interest is in whether managers use and investors treat extreme language as informative signal or not.

Most of prior studies on the role of language in corporate reporting have focused on a single, but important, attribute of language, tone, which captures the extent to which a body of text contains positive or negative words (e.g., Tetlock (2007), Tetlock et al. (2008), Loughran and McDonald (2011), and Huang et al. (2013)). To calculate tone, these studies use popular sentiment orientation dictionaries that classify words into positive and negative categories.[5] Traditional dictionaries, however, do not capture the degree of extremity that each word or phrase exhibits. As a result, it is impossible to test whether market participants respond to extreme language in corporate disclosures using existing positive vs. negative word classifications. Moreover, many prior studies use word weighting functions (e.g., inverse document frequency), likely conflating a word's frequency of occurrence with its linguistic extremity.[6]

To better understand managers' language choices (extreme vs. moderate, positive vs. negative) in earnings conference calls, we create a dictionary of linguistic extremity by ex-

[5] Henry and Leone (2015) review different methodologies to measure linguistic tone.

[6] Section 2.4.2 compares measures of linguistic extremity to popular word weightings in the literature.

tracting all of the adjective, noun, and verb words and phrases from a large sample of 60,940 earnings conference calls. Our final dictionary consists of 23,355 words and phrases. Having created this large and fairly comprehensive corpus, we employ human annotators on Amazon Mechanical Turk to rate these words and phrases in terms of their *signed linguistic extremity*. More specifically, each entry in the dictionary gets rated by multiple individuals, who indicate how positive or negative the word or phrase would be in the context of an earnings announcement on a scale from "extremely negative" (-5) to "extremely positive" (+5). This approach ultimately allows us to distinguish between words with different degrees of both linguistic extremity and tone (e.g., *good* vs. *terrific*; *bad* vs. *terrible*) and, thus, examine whether extremity and tone are the same or different measures of management communications with investors.[7]

For our initial tests of the investor response to extreme language, we focus on abnormal trading volume and abnormal stock returns around earnings conference calls. Stock returns capture the average change in investors' beliefs following the event, while trading volume reflects the differences in investors' reactions to the event (Beaver (1968), Bamber et al. (2011)). Although an association between extreme language and trading volume would not, on its own, imply that investors believe management, *per se*, it does serve as a measure of whether the market views such communication as informative. Controlling for performance and other firm characteristics and time effects, we find that abnormal trading volume is much more strongly associated with extreme language in the conference call than with moderate language. In terms of economic magnitude, one standard deviation increase in linguistic extremity results in a 7.04% increase in the level of abnormal trading volume around the call, whereas moderate words stimulate only a 2.8% more trading per a standard deviation increase in the moderate language. When we decompose linguistic extremity into positive and negative components, we find that both extreme positive and extreme negative language is associated with higher abnormal trading volume, whereas moderate positive and negative language has weak or no association with trading activities. In summary, extreme language in earnings conference calls, whether positive or negative, appears to generate a significant

[7]We validate our MTurk ratings using graduate business students and find that the correlation between responses of these two different groups of people is 94.4%, indicating a strong consistency in word ratings. We provide more details on our dictionary and measures in Sections 2.3 - 2.5.

amount of investor interest and disagreement.

In our returns tests, we find that event-period abnormal returns are positively associated with signed linguistic extremity and that the market reaction to extreme language is much stronger than to moderate language. Specifically, we find that a one standard deviation increase in signed linguistic extremity is associated with abnormal stock returns that are 20.1% larger relative to the median absolute price reaction to the earnings conference call. In contrast, a one standard deviation increase in the amount of moderate positive language relative to moderate negative language results in abnormal stock returns that are only 5.8% larger than the median absolute price reaction. When we split our measures into their positive and negative components, we find strong market reactions to both extreme positive and extreme negative language. Together with the volume results, these returns results paint a picture of investors paying attention to the type of language used in the conference call. Extreme language, in particular, appears to generate considerable disagreement among investors and is associated with greater price response to earnings conference calls, regardless of whether the extreme language is positive or negative. Thus, investors appear to largely treat extreme language as an informative signal as it stimulates significant trading activity and generates stronger event-period price reactions.

To provide further evidence on investors' response to extreme language, we examine whether market reactions documented above depend on a firm's information environment and investor processing costs. Following prior literature on market reactions to earnings announcements (e.g., Chambers and Penman (1984), Bernard and Thomas (1989), and Hirshleifer et al. (2009)), we use firm size, the number of analyst following, and the number of institutional owners as proxies for the relative importance of earnings conference calls to investors. Investors of large firms or firms with larger analyst following will have more sources of information about a company, whereas investors of smaller companies or companies with fewer analysts will likely have to rely more heavily on earnings conference calls as a major source of information. In a similar vein, if we assume that institutional investors are more sophisticated than individual investors at collecting and processing various sources of information, then stock prices of companies with more institutional investors will likely be less sensitive to public information releases, such as earnings conference calls. We find that for our sample all three measures are highly correlated and ultimately proxy for firms' information environment. Consistent with our expectations, we find that the effects of extreme language on trading volume and stock returns are strongest for firms with weaker information environments, where the marginal investor likely has to rely more heavily on public information releases, such as earnings conference calls.

In addition to examining the event period market reactions, we also examine how analysts respond to extreme language. While analysts obtain information relevant to their forecasts from various channels (both private and public), their active participation in conference calls and increased number of subsequent forecast revisions suggest that analysts seem to value information disclosed in conference calls. We use three measures of analyst activity following the earnings calls (i.e., the absolute amount of the forecast revision, the percentage change in the revision, and the proportion of forecasts associated with a forecast upgrade) and find consistent results for each - namely, that analysts react to extreme language in earnings calls more strongly than they do to more moderate language. These findings suggest that analysts, similar to investors, attend to the type of language used in conference calls and, at least on average, find the nuances of language to be informative.

To better understand managers' intent in choosing to use extreme language in earnings conference calls, we next relate extremity of language to one-year-ahead earnings and sales scaled by assets. Consistent with extreme language carrying information about future operating results, we find that extreme language is positively associated with future earnings and sales. In contrast, moderate language exhibits no or weak association with future performance. Moreover, the economic magnitude of the estimated coefficients is more than two times greater for extreme language than for moderate language. When we split our signed extremity variables into their positive and negative components, we find that both positive and negative extreme language is indicative of future earnings and sales. As such, managers appear to, on average, use extreme language to convey information about future operating performance and market reactions to extreme language documented earlier are, at least partially, justified.

We interpret our results so far as the evidence that extremity of language matters, as all of our tests examine the relative importance of extreme vs. moderate language, be it positive or negative. To provide further evidence that linguistic extremity captures an important, but a different attribute of language from positive vs. negative tone, we provide a univariate evidence on the relation between extreme language and the event- and post-event period abnormal returns, conditional on the overall tone of the earnings call and earnings surprise. We observe that market reactions to unexpected earnings strongly depend on the extremity of language in the conference call. For instance, market reactions to low (high) unexpected earnings is about 1.047% (2.274%) higher when language is more extreme and the overall tone of the earnings call is low. Similarly, market reactions to low (high) unexpected earnings is about 2.140% (2.537%) higher when language is more extreme and the overall tone of the earnings call is high. In other words, message tone (positive vs. negative) and message intensity (extreme vs. moderate) capture complementary yet different information in earnings calls. We also see no significant drifts or reversals in prices in the 60-day return window after the earnings conference call, suggesting that investors price the information in earnings conference calls correctly.

As a final test of the information value of extreme language in earnings calls, we test the significance of extreme language while explicitly controlling for its tone. Specifically, we first calculate traditional measures of positive and negative tone (i.e., proportions of positive and negative words in the conference call) and then measure the extent of extremity in positive (negative) language as the number of extreme positive (negative) words divided by total positive (negative) words in the conference call. Measures of extremity constructed in this manner are different from tone and capture the extent to which tonal words (positive and negative) are extreme. Controlling for positive and negative tone, firm characteristics and time effects, we find that measures of linguistic extremity have an incremental explanatory power in all of our market reactions, analyst revisions, and future performance tests. Overall, these results suggest that investors price information in language choices of management correctly with extreme language carrying new information to the market.

Our paper builds on and contributes to the emerging literature on linguistic analysis in corporate reporting. Many papers on corporate disclosures use binary measures of tone to examine how the relative tone of language that accompanies financial information influences market participants and facilitates price discovery (e.g., Tetlock (2007), Tetlock et al. (2008), Feldman et al. (2009), Loughran and McDonald (2011), Price et al. (2012), Huang et al. (2013), Jegadeesh and Wu (2013), and Blau et al. (2015)).[8] For instance, Price et al. (2012) find significant price reactions to tone in earnings conference calls. We extend their work by showing that, in addition to tone, investors respond to extreme language in earnings conference calls. We also document new results on firms' information environments and market reactions to conference calls, analyst forecasting activities following the calls, and the information content of extreme language for future operating performance. Recognizing that investors react differently to different words in corporate disclosures, Jegadeesh and Wu (2013) estimate price reactions to individual positive and negative words to infer each word's explanatory power for abnormal returns, which they call "word power." We contribute to this work by showing that tone and extremity capture complementary yet different information in earnings calls and that our measures of linguistic extremity are different from "word power" weights in Jegadeesh and Wu (2013) or traditional word frequency weightings in prior literature (see Section 2.4.2).

Our paper also contributes more generally to research on vivid and extreme language. Many experimental studies examine how individuals react to extreme and vivid language (e.g., Burgoon et al. (1975), Nisbett and Ross (1980), Aune and Kikuchi (1993), and, more recently, Hales et al. (2011)). Recent archival studies on corporate executives use linguistic analysis to study CEO and CFO behavior. Blankespoor and DeHaan (2015) examine CEO quotes in media coverage and find that the clarity and vividness of what CEOs say can impact CEO visibility and career outcomes. Relatedly, Larcker and Zakolyukina (2012) develop a machine-learning model to predict deceptive vs. truthful communication of executives, where deceptive communication is the one that results in future restatements of financial statements. They analyze different linguistic characteristics of management speech and find that deceptive CFOs and CEOs use more references to general knowledge, fewer non-extreme positive emotion words, and fewer references to shareholder value. We extend this work by examining whether extreme language is meaningful to investors and analysts and whether it

[8] A relatively smaller number of papers examines linguistic attributes other than tone, such as readability and complexity (Li (2008), You and Zhang (2009), Miller (2010), Rennekamp (2012), Bonsall and Miller (2017), Chychyla et al. (2018)), content and similarity (Hanley and Hoberg (2010), Hoberg and Maksimovic (2015)), spontaneity (Lee (2015)), and linguistic formality (Rennekamp and Witz, 2017).

carries information about future operating performance.

Taken together, we contribute to the literature by introducing the notion of linguistic extremity. Specifically, we establish a dictionary that can be used to measure both tone and extremity of language in corporate reporting. We show that linguistic extremity captures an important, yet different from tone, attribute of language. Moreover, by showing that the market responds differently to extreme language than to moderate language, we provide additional evidence that the effects of linguistic tone documented in prior research are likely driven by the investors' paying attention to the words that managers choose, as opposed to being driven by a correlated omitted variable.

The rest of the paper proceeds as follows. In Section 2, we summarize related literature and describe our conference call sample, dictionary creation, and measurements of linguistic extremity. In Section 3, we present our main empirical results. We conclude in Section 4.

2 Data and Methodology

2.1 Related Research

In recent years, many studies in accounting and finance have focused on the analysis of firms' qualitative disclosures. The current consensus in the literature is that qualitative disclosures are incrementally informative above and beyond traditional financial factors. For instance, Price et al. (2012) measure the tone of earnings conference calls and find that it is significantly associated with market reactions to such events. Blau et al. (2015) calculate abnormal tone as the difference between the tone of the introduction and the tone of the question and answer section of the conference call. They find that short sellers interpret abnormal tone differently than naive investors. Milian and Smith (2017) find the amount of praise by analysts on an earnings conference call is strongly associated with market reactions to the call. Larcker and Zakolyukina (2012) develop a machine-learning model to predict deceptive vs. truthful communication of executives. They analyze different linguistic characteristics of management speech and find that deceptive CFOs and CEOs use more references to general knowledge, fewer non-extreme positive emotion words, and fewer references to shareholder value.[9]

All these studies use binary measures of tone (positive vs. negative) to examine the usefulness of qualitative disclosures to market participants. Texts that use more positive words than negative words are classified as relatively more optimistic and those that use more negative than positive words are considered relatively pessimistic. While counting positive and negative words in texts is intuitive and easy to apply, it does not capture other aspects of communication, such as linguistic extremity. For example, with the binary approach, words like 'good' and 'superior' or 'problem' and 'disaster' or 'bad' and 'terrible' are each treated equally and coded as 1 (if a word is positive) or -1 (if a word is negative), ignoring potential differences in a word's linguistic extremity.

In this study, we contribute to and extend prior literature on tone of corporate disclosures by developing a dictionary of linguistic extremity and empirically examining the information value of extreme language for the market. We aim to demonstrate that market participants respond to extreme words differently that to more moderate words, be those words positive or negative. Additionally, we aim to understand whether managers use extreme words to inform or delude the market.

2.2 Earnings Conference Calls Sample

Since our intent is to analyze how investors respond to linguistic extremity, we begin by building a dictionary of words and phrases used in earnings conference calls. Earnings conference calls provide us with a unique setting for analyzing the use and impact of extreme (spoken) language on investors' judgments.[10] When compared to the SEC disclosures, earnings calls are more timely, less formal, easier to follow and understand, and less sanitized

[9] There is also a large stream of literature analyzing linguistic properties of mandatory SEC disclosures. For instance, Li (2010a) documents that the tone of forward-looking statements in the MD&A section of 10-K and 10-Q reports is positively associated with future earnings. Feldman et al. (2009) find a significant association between tone change in the SEC filings and short window market reactions around filing dates. Similarly, Loughran and McDonald (2011) use SEC filings to develop domain specific word lists and show that their word lists are associated with market reactions around 10-K filing dates and unexpected earnings.

[10] Generally, the conference call starts with a brief introduction of the management team present on the call and a legal disclaimer about forward-looking statements. Then one or more of company executives (usually the CEO, CFO, or both) discuss the operating performance for the quarter just ended and provide information on the company's prospects, plans and operations. After these introductory statements by top management, the call is opened for questions from participating analysts and investors.

by lawyers.[11] For example, executives of Best Buy Co. (BBY) in the third quarter of fiscal 2006 earnings conference call were making the following statements:

> The good news was that we had a strong start to the holiday season [...]. We've continued to grow our market share. We generated a respectable comparable store sales [...]. We also continued to see significant improvement in our gross profit rate. These variables we consider are very encouraging. We had very ambitious plans for the quarter with strong growth goals and transformation goals [...]. We made a tremendous number of investments in our portfolio of capabilities [...]. We saw exciting top-line results from the segmented stores [...] I was very pleased with our revenue results [...] and I'm confident that a competitive offering and the relationships that we've built with our customers during the year drove our strong performance. Our results give me optimism for the fourth quarter [...]. We're encouraged by our progress with the transformation.

In contrast, Best Buy's Management Discussion and Analysis section in the follow-up 10-Q report looks sterile and linguistically pallid relative to the conference call. 'Strong' and 'increase' are the most positive words used in the document. Words and phrases like 'exciting', 'very encouraging', 'very ambitious', 'tremendous', and 'very pleased' do not appear in the 10-Q report. This example provides an illustration of significant differences in language of written SEC disclosures and spoken earnings conference calls.

To construct our sample of earnings conference calls, we turn to *www.seekingalpha.com*. Seeking Alpha is one of the largest investor-oriented websites in the United States that covers a broad range of publicly traded companies and provides a free access to earnings conference call transcripts. The first step was, therefore, to write a computer program to download in HTML format all transcripts of earnings calls available on Seeking Alpha for the years 2006 to 2015 and to extract the textual content from each of these files.[12]

Next, we attempted to find matching COMPUSTAT data for each conference call. Each transcript contains an identifying information about a company including company name, ticker, and the date of the earnings call, which can be matched to tickers and earnings announcement dates from COMPUSTAT.[13] The Exchange Act Form 8-K (Section 206) states

[11] Annual and quarterly SEC disclosures are widely criticized for being uninformative, redundant, overwhelmingly long, and polished by lawyers (Monga and Chasan, 2015).

[12] Seeking Alpha was founded in 2004, but a comprehensive coverage of firms on the website started in 2006. Chen et al. (2014) is one of the first large-scale studies that uses Seeking Alpha's articles to study investor opinions, market returns, and earnings surprises.

[13] To ensure the accuracy of the matching, we performed extensive manual checks of matched company

that conference calls that are made publicly available and occur within 48 hours of the earlier press release will not trigger additional 8-K disclosures. Not surprisingly, most companies in our sample hold earnings calls on the day of the earnings announcement (around 80%) or on the following day (around 18%), and a few companies hold the call within one week of the earnings announcement (around 2%). From our initial sample of 60,940 earnings conference calls, we were able to obtain matching COMPUSTAT data for 45,056 firm-quarters.

We then proceeded to download financial statements, analyst forecasts, and market data from COMPUSTAT, IBES, and CRSP. For each firm-quarter, we required a non-missing values for the event-period abnormal return and abnormal trading volume, at least one analyst forecast, the number of analysts following the firm, and enough information to calculate return-on-assets, accruals, future earnings and sales, pre-announcement return, market-to-book ratio, leverage, Altman's Z-Score, earnings volatility, return volatility, firm age, and number of business and geographic segments. To estimate earnings surprise, we used the most recent analyst consensus forecast of one- or two-quarters-ahead earnings issued or reviewed in the last 60 days before the earnings announcement. We also required at least 1,000 words in each earnings conference call transcript.[14] These data requirements further reduced our sample to 35,155 firm-quarter observations.

The event-period cumulative abnormal volume around the earnings conference call is calculated as the logged difference between the announcement period trading volume and expected trading volume following the methodology in Campbell and Wasley (1996) ($CAV[0,2]$). The event-period cumulative abnormal return around the earnings conference call is calculated as the difference between the buy-and-hold return of the firm and that of a size, book-to-market, and momentum matching portfolio over the 3-day event window ($BHAR[0,2]$). Table II outlines all variables with definitions and data sources used in our analyses.

names and earnings announcement dates.

[14]Sometimes Seeking Alpha publishes a short summary of an earnings call instead of the whole transcript. We use a 1,000 words cutoff to ensure that we capture the whole transcript. We note, however, our results are similar if we do not impose this requirement.

2.3 Measurement of Linguistic Extremity

There are many experimental studies in psychology and communication that analyze the effects of language intensity, extremity, and vividness on people's perceptions and decision making (see for e.g., Nisbett and Ross (1980), Aune and Kikuchi (1993), Hamilton (1998), Andersen and Blackburn (2004), Clementson et al. (2014)). In the context of financial reporting, Hales et al. (2011) develop an experiment to analyze the impact of vivid language on investors' judgments and find evidence that investors react to differences in language, particularly when the underlying information is inconsistent with their preferences.

To measure the linguistic extremity of earnings conference calls, we relied on the same theories as Aune and Kikuchi (1993) and Hales et al. (2011) and adopted a methodology developed in Taboada et al. (2011) for measuring the sentiment of reviews of books, movies, hotels, etc. First, we extracted all adjectives, nouns, and verbs that occurred in more than 1% of all earnings conference calls (60,940 earnings call transcripts).[15] We then deleted finance and accounting terms from Brindley and Law (2008) and Law (2010) as well as all stop words, names, and generic terms. We then created the union of our word lists and Loughran and McDonald (2011)'s positive and negative word lists. For each word in the merged dictionary, we also tried to find synonym words and phrases using the Microsoft Word's thesaurus feature. In this manner, we built a comprehensive list of words and phrases (not limited to words and phrases in our sample of conference calls) that could be used in other settings, such as other types of corporate disclosures or out-of-sample tests of future conference calls. Our final dictionary (*DICT*) consists of 23,355 words and phrases, where 6,395 are adjectives, 8,361 are nouns, 2,363 are verbs, 187 are adverbs, and 6,049 multi-word phrases.[16]

The next step was to measure the linguistic extremity of each word in *DICT*. To do that,

[15] We used all available conference calls to construct our dictionary. This enabled us to have a comprehensive (i.e., not limited to a specific sample) coverage of words and phrases.

[16] *DICT* contains many words that were not identified by Loughran and McDonald (2011) as they never or very rarely occur in regulated SEC filings. For example, positive words 'phenomenal' and 'terrific' occur in about 5% and 17% of earnings calls observations, respectively, but are not used in 10-K reports. Similarly, negative words like 'awful' and 'terrible' are used in about 3% and 6% of earnings calls observations, respectively, but very rarely occur in the SEC filings.

we employed individuals using Amazon's Mechanical Turk service (MTurk).[17,18] MTurk is a marketplace for small-scale tasks that require human intelligence. It is becoming a widely-used resource among researchers for various tasks including experimental studies, subjectivity and sentiment analysis (see Paolacci et al. (2010), Buhrmester et al. (2011), Blankespoor, Hendricks, and Miller (Blankespoor et al.)). Briefly, MTurk connects people (Requesters), who have tasks that require human intelligence, with people (Workers), who can perform such tasks, usually for relatively little pay. Quality of the completed task is mostly controlled through workers' approval ratings for previously completed tasks. In addition, requesters can reject work that was done incorrectly, in which case workers are not paid and their approval ratings go down.

Our human intelligence task (HIT) on MTurk consisted of rating 50 randomly selected words and phrases from *DICT* on a scale ranging from "-5" for extremely negative to "+5" for extremely positive, where "0" indicates a neutral word or phrase. Each task took on average 5 minutes to complete. We employed the most highly qualified English speaking workers on MTurk (Masters) and paid them \$0.45 for successfully completing the task.[19] To ensure the high quality of responses, we randomly inserted meaningless entries in each HIT and monitored whether workers were able to identify those entries. Workers had to decide how positive or negative each word or phrase is based on its meaning in the context of earnings announcements. Figure 1 shows a snippet of our HIT on MTurk's web page. For each entry in our dictionary, we required 5 ratings, which gave us enough observations to check the quality of responses. In addition to word ratings, we collected basic demographic information on each worker's educational background, gender, and nationality.[20]

Using this method, we obtained a set of linguistic extremity ratings for all 23,355 words

[17] See *www.mturk.com*.

[18] IRB approval for this portion of our study was obtained before having individuals rate the words and phrases in our dictionary.

[19] According to MTurk, Masters is an elite group of workers who have demonstrated high accuracy on specific types of HITs. Workers achieve a Masters distinction by consistently completing HITs with a high degree of accuracy across a variety of requesters. Masters are continuously monitored to remain Mechanical Turk Masters. The compensation of \$0.45 per HIT was determined based on a review of existing HITs on MTurk that required similar amount of time and effort.

[20] Based on a self-reported demographic information, the average age of workers is 37, the majority of workers (around 80%) are from the United States, around 54% have an undergraduate or graduate degree, and around 52% are female.

and phrases in *DICT*. Rating responses for each dictionary entry were reasonably consistent. However, even though our HIT was designed to discourage random responses from workers, we also took additional steps to minimize the influence of inattentive responses or misinterpretations of words. For example, if a word was rated "-4", "-3", "-4", and "-5" by four raters, and rated "+3" by the fifth rater, we ignored the divergent response of the fifth rater and used the other four ratings to determine the word's rating. Those words that workers indicated they were unable to rate were excluded from the dictionary.[21]

2.4 Linguistic Extremity and Frequency

2.4.1 Distribution of Extremity Rankings

Figure 2 shows the frequency distribution of all responses by their signed linguistic extremity. Consistent with intuition, we find that extremely positive and extremely negative words account for a smaller proportion of words in our dictionary, whereas moderately positive and negative words account for the majority of entries in *DICT*. The distribution appears to be roughly symmetric in that the frequency of words in the dictionary is declining in absolute extremity rating.

Given this roughly normal distribution, many of the words and phrases we extracted from conference calls were rated as essentially neutral. More specifically, 2,105 adjectives, 4,492 nouns, 993 verbs, and 2,252 multi-word phrases (or around 42% of all words) received an absolute average rating between 0 and 1. By way of example, the words and phrases 'stay close', 'digestive', 'aggregation', and 'put in writing' all received an absolute average rating of less than 1. Because such words do not carry any tonal information, we removed them from our dictionary. Our final rated dictionary (*RATED_DICT*), therefore, consists of 13,513 words and phrases with average ratings between either -5 and -1 or between 1 and 5.

Table I provides examples of words and phrases in each rating category and extracts from earnings conference calls that use these words and phrases. Overall, ratings in *RATED_DICT* seem to be consistent with our intuition. For example, the words 'exceptional', 'excellence',

[21] We also selected a random set of 200 words and asked graduate students in business to rate them, requiring 10 ratings per each word. The correlation between ratings obtained through MTurk and those by graduate students is 94.4%, indicating a strong consistency in responses across different groups of people.

and 'amazing' are rated as extremely positive, whereas the words 'default', 'terrible', and 'devastate' are rated as extremely negative. Similarly, the phrases 'extremely well' and 'very poor' are rated as extremely positive and extremely negative, respectively. In contrast, the words 'steady', 'produce', and 'sufficient' are rated as moderately positive, whereas 'limitation', 'unexpected', and 'complexity' are rated as moderately negative.

2.4.2 Extremity Rankings and Popular Word Weightings

In light of Figure 2, it is possible that extreme words are simply a proxy for infrequent words, i.e., words that occur rarely in earnings conference calls. To examine the relation between extremity ratings and word usage in earnings calls, we calculate a popular *inverse document frequency (idf)* metric for every single-word entry in *RATED_DICT*.[22] We follow Jurafsky and James (2000) and define an inverse document frequency, *idf*, for word i as:

$$idf = log(\frac{N}{df_i}),$$

where N denotes the number of earnings call transcripts in the sample and df_i denotes the number of transcripts that contain at least one occurrence of word i. Intuitively, the inverse document frequency of a rare word is high, whereas the inverse document frequency of a frequent word is low.

In untabulated tests, we find that the correlation between absolute extremity ratings of single words in *RATED_DICT* and their *idf* values is 4.8%, indicating only a weak association between words' frequencies in earnings conference calls and their extremity rankings. In addition, the coefficient of variation of *idf* stays at around 40% across all extremity groups. Said differently, there are common and uncommon words with different degrees of extremity. For example, the words 'miraculous', 'utopia', 'marvelously', 'tyrannical', 'disgraceful', and 'terrorize' are rated as extreme, but have *idf* values greater than 7, indicating their infrequent usage in earnings calls. In contrast, the words 'wonderful', 'terrific', 'excellent', failure', and 'awful' are also rated as extreme, but are much more common in conference calls, having *idf*

[22] We use single words to calculate inverse document frequencies since multi-word phrases will have higher inverse document frequencies by construction. For example, a phrase 'gain recognition' has an *idf* value of 7.05 indicating its infrequent usage in earnings conference calls, whereas words 'gain' and 'recognition' have *idf* values of 0.24 and 1.75, respectively, indicating a frequent usage of these single words in earnings calls.

values of less than 3.

Recognizing that words likely have variation in their usage and meaning, Jegadeesh and Wu (2013) introduce an alternative to *idf* weighting scheme, which relies on market reactions to individual words. Specifically, they run a regression model of the 10-K filing period returns on all positive and negative word counts in each 10-K to estimate the strength of individual words (referred to as "word power weights") in explaining market reactions to 10-Ks. It is possible that the explanatory "power" of each word in market reaction models is driven by each word's extremity ranking which we capture in *RATED_DICT*. However, there might be a number of other factors contributing to the explanatory "power" of each word. Moreover, "word power weights" likely change with time periods and samples of firms used to estimate the statistical models, whereas our rankings are invariant and intended to capture linguistic extremity of individual words and phrases.

To test the similarity between "word power" weights and our extremity rankings, we replicate the methodology in Jegadeesh and Wu (2013) in our setting. Specifically, we run a regression of abnormal returns around earnings conference calls on individual word counts using Loughran and McDonald (2011)'s dictionary and then compare the estimated word power weights and extremity rankings of those words. We find that the correlation between our extremity rankings and word power weights is around 2% (statistically insignificant). Moreover, we find that word power weights change over time and with sub-samples of firms used to estimate the statistical relations. Therefore, it is unlikely that the methodology developed in Jegadeesh and Wu (2013) captures stable properties of language, such as linguistic extremity that we study.

2.5 Variables Measuring Extreme and Moderate Language in Earnings Conference Calls

The extremity ratings in *RATED_DICT* provide a new dimension for measuring document information content. However, because our intent is to compare how market participants react to extreme vs. moderate language, we construct separate variables for the proportion of words in a conference call falling into each category. Doing so allows us to estimate separate regression coefficients for each type of language and compare them. Using rounded ratings of words and phrases in *RATED_DICT* as the basis for word counts in earnings conference calls, we create four measures of linguistic extremity as follows:

1. $TotalExtreme = \frac{\text{Number of Words with Absolute Ratings of "4" or "5"}}{\text{Number of All Words}}$,

   where *TotalExtreme* measures the proportion of all extreme words and phrases in a given earnings conference call.

2. $TotalModerate = \frac{\text{Number of Words with Absolute Ratings of "1", "2", or "3"}}{\text{Number of All Words}}$,

   where *TotalModerate* measures the proportion of all moderate words and phrases in a given earnings conference call.

3. $SignedExtreme = \frac{\text{Number of Words Rated as "4" or "5"} - \text{Number of Words Rated as "-4" or "-5"}}{\text{Number of All Words}}$,

   where *SignedExtreme* measures the relative usage of extreme positive versus extreme negative words in a given earnings conference call.

4. $SignedModerate = \frac{\text{Number of Words Rated as "1", "2", or "3"} - \text{Number of Words Rated as "-1", "-2", or "-3"}}{\text{Number of All Words}}$,

   where *SignedModerate* measures the relative usage of moderate positive versus moderate negative words in a given earnings conference call.

In addition, we further decompose our linguistic extremity measures into their positive and negative components to allow for (potentially nonlinear) differences in how markets respond to extreme and moderate language when that language is positive rather than negative.[23]

The above linguistic extremity measures are conceptually similar to traditional measures of *Tone* (scaled difference between positive and negative words) since they combine the positive and negative tone and extremity of language. To examine whether linguistic extremity is informative by itself (i.e., beyond *Tone*), we separate the concept of extremity from the concept of tone. Specifically, we measure the extent of extremity in positive (negative) language as the number of extreme positive (negative) words divided by total positive (negative)

[23]One important aspect in measuring linguistic extremity is the occurrence of negators in text (e.g., not great, not bad, nothing spectacular, not well). In untabulated analyses, we follow a methodology in Taboada et al. (2011) to account for negations of positive and negative words and phrases that occur in earnings call transcripts. Specifically, we split every sentence in an earnings call into clauses and check for the presence of negators (i.e., not, never, no, none, etc.). If negator is present in a given clause, we shift an assigned extremity rating in the opposite direction by 4. That is, "Our performance was not *bad*", initially rated as -3, is now assigned a new rating of 1 (-3 + 4); "Our performance was not *great*", initially rated as +5, is now assigned a new rating of 1 (5 - 4). The correlation between our main measures and those that account for negation is around 90%. All our inferences remain unchanged when we use these measures in our analyses.

words, *ExtrWordsInPositive* (*ExtrWordsInNegative*). Measures of extremity constructed in this manner are not collinear with *Tone*, but rather are more likely to capture the extent to which tonal words (positive and negative) are extreme. Table II provides formal variable definitions and measurements.

2.6 Descriptive Statistics

Table III presents the descriptive statistics for our measures of linguistic extremity and other quantitative variables related to the 35,155 earnings announcements accompanied by earnings conference calls over the period 2006 to 2015. The mean (median) of *PosExtreme* is 0.74% (0.72%) and of *PosModerate* is 15.99% (15.98%). The mean (median) of *NegExtreme* is 0.10% (0.08%) and of *NegModerate* is 2.89% (2.85%). When we combine extreme and moderate, positive and negative scores into aggregate measures, the mean (median) of *SignedExtreme* is 0.64% (0.62%) and the mean (median) of *SignedModerate* is 16.72% (16.72%). Further, the mean (median) of *ExtrWordsInPositive* and *ExtrWordsInNegative* is 4.42% (4.31%) and 3.22% (2.84%), respectively, indicating that a portion of tonal language is extreme. Consistent with Price et al. (2012), this descriptive evidence also suggests that managers tend to use more positive than negative words in earnings conference calls. This tendency is in sharp contrast with the word usage in the SEC filings, where the majority of tonal words are negative (see Table II in Loughran and McDonald (2011)).

In Table IV, we present unconditional Pearson correlations for our main variables of interest. We find that event period abnormal trading volume and abnormal stock returns are more strongly correlated with measures of extreme language than with measures of moderate language. Similarly, analyst forecast revisions following earnings conference calls and future performance measures are more strongly correlated with extreme language than with moderate language. Similarly, we observe strong correlations between *ExtrWordsInPositive*, *ExtrWordsInNegative* and our main dependent variables of interest. In sum, these univariate results suggest that extremity of language carries new information to the market.

3 Results

This section summarizes our empirical findings. In Section 3.1, we examine the association between linguistic extremity in conference calls and event period abnormal trading volume and stock returns. In Section 3.2, we analyze whether a firm's information environment and information processing costs moderate investors' reactions to linguistic extremity. Section 3.3 studies an association between analysts' forecast revisions following earnings conference calls and extreme and moderate language used in conference calls. Section 3.4 examines whether linguistic extremity is informative about future operating performance. Finally, Section 3.5 presents results with alternative measures of linguistic extremity.

3.1 Market Reactions to Extreme Language in Conference Calls

We start by examining the impact of linguistic extremity on investors' reactions to earnings conference calls. Compared to written disclosures filed with the SEC, which tend to be formal and heavily influenced by corporate legal departments, language in conference calls is less restricted and allows managers to express both their personality and their beliefs about current and future performance. If investors not only pay attention to nuances in the language used by management, but also find management's statements persuasive and credible, even when extreme, then we should see stronger market reactions, in terms of trading volume and stock returns, in response to extreme language than to moderate language. Importantly, if extreme language is just a firm or a manager fixed effect and does not carry new information, then we should observe no volume or price reactions to extremity as firm- or manager-specific styles are known to the market.

To test our predictions, we regress cumulative abnormal volume (relative to expectation) and cumulative buy-and-hold abnormal return (relative to return of size, book-to-market, and momentum matching portfolio) over the 3-day event window on extreme and moderate scores and a set of control variables:

$$CAV[0,2] = \alpha_0 + \alpha_1 TotalExtreme + \alpha_2 TotalModerate + A \times Controls_v + \varepsilon, \qquad (1)$$

$$BHAR[0,2] = \beta_0 + \beta_1 SignedExtreme + \beta_2 SignedModerate + B \times Controls_r + \varepsilon, \quad (2)$$

where $TotalExtreme$ and $TotalModerate$ are the sum of extreme positive and extreme negative scores and moderate positive and moderate negative scores, respectively; $SignedExtreme$ and $SignedModerate$ are the difference of extreme positive and extreme negative scores and moderate positive and moderate negative scores, respectively; $Controls_v$ include unexpected earnings, loss, loss interacted with the unexpected earnings, indicator variables for high and low unexpected earnings, the absolute event period abnormal return, return-on-assets, accruals, earnings volatility, market-to-book, leverage, Altman's Z-score, stock returns, return volatility, analyst following, number of business and geographic segments, firm age, financial industry indicator, forward-looking disclosures in earnings calls, risk and uncertainty disclosures in earnings calls, and year-quarter fixed effects; $Controls_r$ are the same as $Controls_v$, but include (exclude) pre-announcement return (absolute event-period abnormal return) as a control variable.[24,25]

To examine whether investors react more strongly to extreme language in the conference call, we use an F-test to test the equality of coefficients on extreme and moderate language. Specifically, we test whether $\alpha_1 > \alpha_2$ in Eq. (1) and $\beta_1 > \beta_2$ in Eq. (2). The first (last) four columns in Table V provide coefficient estimates for the association between abnormal trading volume, $CAV[0,2]$, and proportions of extreme and moderate words used in earnings conference calls, excluding (including) control variables. We suppress the coefficient estimates for all control variables for exposition purposes. We find that abnormal trading volume is much more strongly associated with extreme language in the conference call than with moderate language. Specifically, a one standard deviation increase in the proportion of extreme words in the conference call is associated with an 7.04% increase in abnormal trading volume (0.0025×28.171), whereas trading volume increases by 2.8% per a standard deviation increase in the moderate language (0.0129×2.204). Equality of these two coefficients ($\alpha_1 = \alpha_2$) is strongly rejected (F = 27.35). To put the magnitude of this effect into context, we

[24] Control variables are selected following related work by Tetlock (2007), Tetlock et al. (2008), Li (2010a), Rogers et al. (2011), Price et al. (2012), Davis et al. (2012), Huang et al. (2013), Henry and Leone (2015), and Bonsall and Miller (2017).

[25] Our results are not sensitive to whether we include industry- or firm- fixed effects or to whether we calculate unexpected earnings relative to the same quarter last year.

look at the impact of moving from the lowest extremity decile to the highest extremity decile and find that abnormal trading volume is, on average, 24% higher for firms in the highest extremity decile.

Next, because tone could influence how investors react to linguistic extremity, we separately examine extreme and moderate scores for positive and negative language. As reported in Column (8) of Table V, we find that the association with abnormal trading volume remains stronger for both positive and negative extreme scores relative to positive and negative moderate scores. Overall, given that our measures in Table V are constructed using individual words' and phrases' extremity rankings, we interpret these results as evidence that investors respond more strongly to extreme than to moderate language in earnings conference calls.

Similar to results in Table V, we find that extreme language in the conference call is more strongly associated with the event-period cumulative abnormal return ($BHAR$) than moderate language. The first (last) four columns in Table VI provide coefficient estimates for the association between $BHAR[0,2]$ and $SignedExtreme$ and $SignedModerate$, excluding (including) control variables. We find that, while both extreme and moderate language are significantly positively associated with abnormal returns, linguistic extremity is associated with a stronger price reaction. One standard deviation in $SignedExtreme$ results in a 0.87% (0.0026×3.337) higher abnormal return, while a comparable increase for $SignedModerate$ results in a 0.25% (0.0172×0.147) increase in return (both significant at the 1% level). Equality of these two coefficients ($\beta_1 = \beta_2$) is strongly rejected (F = 235.39). To place these results in context, the median absolute price reaction around earnings conference calls in our sample is 4.33%, so a 0.87% (0.25%) increase in the abnormal stock return corresponds to a 20.1% (5.8%) larger price reaction relative to the median price reaction. In addition, if we look at the coefficient of variation of $BHAR$ regressions, we see that the adjusted R^2 of a regression model that uses $SignedExtreme$ as one of the explanatory variables is around 7% higher than the R^2 of a similar model that uses $SignedModerate$ as an explanatory variable.

Next, we decompose $SignedExtreme$ and $SignedModerate$ intro their positive and negative components. Column (8) of Table VI reports the results. We find that the equality of the extreme and moderate coefficients is strongly rejected for both positive and negative language (F = 190.15 and F = 37.3, respectively). These results suggest that investors respond more strongly to extreme than to moderate language, be it positive or negative.[26]

To further explore return reactions to linguistic extremity of earnings conference calls, in Figure 3, we plot $BHAR[0,2]$ against deciles of unexpected earnings and deciles of $SignedExtreme$. The three dimensional surface plot demonstrates that return reaction around earnings calls increases both in the degree of unexpected earnings and linguistic extremity. Firms with high unexpected earnings (decile 10) and high signed extremity (decile 10) experience around 8% event-period cumulative abnormal return, whereas firms with high unexpected earnings (decile 10) and low signed extremity (decile 1) have around 1.7% of abnormal return. This graph helps to rule out the possibility that our results are driven by outliers or other data frictions. Moreover, in untabulated tests, we run all our analyses with decile regressions and get similar results. Taken together, given that our measures capture the extent of extremity in earnings conference calls, we interpret these results as evidence that investors pay attention to the type of language in earnings conference calls and react more strongly to extreme words than to moderate words.

Interestingly, in Tables V and VI, we find that both positive and negative extreme words in the conference call are informative to investors. This finding is different from Loughran and McDonald (2011), who find that only negative words in 10-K prompt significant market reactions. There are several possible explanations for this result. First, earnings conference calls are spoken rather than written information sources and their linguistic characteristics are likely different (e.g., language is less formal). Second, companies may remove or tone down most of the positive statements in their regulated disclosures to reduce their litigation exposure. In contrast, while conference calls are also subject to litigation risk, they are less likely to be sanitized by corporate lawyers because a large portion of the conference call is spontaneous (i.e., driven by analysts' questions). To explore this possibility more, we decompose the content of earnings conference calls into Prepared Remarks and Q&A sections. In the Internet Appendix, we tabulate returns and volume tests for both sections of earnings calls. Similar to our earlier findings, we find stronger market reactions to extreme

[26]For both returns and trading volume specifications, we get similar results if we exclude 2008-2009 financial crisis period from our sample, or if we exclude firms in the financial sector. Further, we get similar results when we limit the content of the earnings conference call to sentences that talk about performance and calculate all extremity measures using such sentences as an input.

language than to moderate language both in the Prepared Remarks and in the Q&A sections of earnings calls. In addition, when we separately measure linguistic extremity of managers' and analysts' statements in the Q&A section, we find that investor reaction to the Q&A section stems from both parties.

3.2 Firm Information Environment and Market Reaction to Extreme Language

Investors' response to the extreme and moderate language in conference calls is likely influenced by firms' information environments and investors' information processing costs. Following prior literature on market reactions to earnings announcements (e.g., Chambers and Penman (1984), Bernard and Thomas (1989), and Hirshleifer et al. (2009)), we use firm size, analyst following, and the number of institutional investors to test the moderating effect of information environment and processing costs on investors' response to extreme and moderate language in conference calls. In our sample, all three measures are highly correlated and ultimately serve as a proxy for the relative importance of earnings conference calls to investors.

The amount of information available about a firm increases in firm size. Larger firms face greater regulatory disclosure requirements and appear more in the financial press and on social media. Therefore, it is likely that earnings conference calls are, on average, less informative for large firms than for small firms. Similarly, analysts have access to information sources other than earnings conference calls (e.g., industry reports, news articles, independent research, etc.), and their forecasts aggregate these various sources of public and private information (Clement et al., 2011). Therefore, investors of heavily followed firms likely react less strongly to information disclosed in earnings conference calls as compared to those firms that have low analyst following. Finally, the degree of institutional ownership is often used as a measure of investor sophistication with the assumption that institutions trade more on private than public information and have lower information processing costs. Compared to retail investors, institutional investors are likely in a better position to interpret language choices of management.[27] By conditioning on the information environment, we can test whether retail investors tend to take management's words at face value and whether information environment moderates the overall market reaction to linguistic extremity.

Each quarter we sort firms into three groups based on their size (small, medium, and large), analyst following (low, medium, and high), and the number of institutional owners (low, medium, and high). We then create three indicator variables for small size, low analyst following, and low institutional ownership firms and interact these variables with *TotalExtreme* and *TotalModerate* in the abnormal trading volume regressions and with *SignedExtreme* and *SignedModerate* in the abnormal stock returns regressions, controlling for firm characteristics and time effects (as in Tables V and VI).

Tables VII and VIII report the results of our cross-sectional sorts for $CAV[0,2]$ and $BHAR[0,2]$, respectively. We find that all our results reported in Tables V and VI hold for largeer firms, firms with more analyst following, and for firms with many institutional owners. However, investors of small firms, firms with fewer institutional owners, and firms with smaller analyst followings tend to react more strongly to extreme words (all coefficient estimates are positive and significant). We do not find consistent differential reactions to moderate words.[28] Overall, these results provide additional support that extreme language is informative to investors, and even more so for firms with weaker information environments and higher information processing costs.

3.3 Analysts' Reactions to Extreme Language in Conference Calls

Another way to test whether markets respond to extreme language in earnings conference calls is to examine the impact that such language has on analyst behavior. While analysts likely use multiple channels (both private and public) to obtain information relevant to the companies they follow, their active participation in earnings conference calls and tendency to

[27]Research in psychology suggests that the natural tendency of individuals when hearing new information is to accept that information at face value, whereas disbelief or skepticism requires effortful information processing and knowledge (e.g., Gilbert (1991), Gilbert et al. (1993), Chen and Chaiken (1999), Petty and Wegener (1999)).

[28]For both volume and returns, we get similar results if we estimate our models on sub-samples of small and large firms, low and high following, or low and high ownership and then test the equality of the estimated coefficients across sub-samples.

subsequently revise their forecasts suggest that analysts are likely well-attuned to information disclosed in conference calls.[29] Therefore, we study the impact of linguistic extremity on analyst forecast revisions made within 10 days following the earnings call.

We use three different measures to capture analysts' response to language in conference calls. Controlling for underlying firm performance and other characteristics (as in Table VI), we first examine how extreme and moderate language influence the magnitude of analyst forecast revisions scaled by price (*AmountOfRevision*). We then test an association between the relative analyst forecast revision (*PercRevision*, i.e., forecast revision in the % terms) and extreme and moderate language. Finally, we test whether the type of language used in conference calls is related to the proportion of analysts revising their forecasts upwards (*PropRevUp*) withing 10 days following the earnings call. If analysts, similar to investors, revise their expectations based on the type of language in earnings conference calls, we expect analyst revision activities to be more strongly associated with *SignedExtreme* than with *SignedModerate*.

In Table IX, we find that linguistic extremity is strongly associated with all three measures of analysts' forecast revisions issued within 10 days after the earnings call. For instance, when we use the amount of forecast revision scaled by price as the dependent variable, the coefficient estimates on *SignedExtreme* and *SignedModerate* are 0.175 and 0.008, respectively, and the equality of these two coefficients is strongly rejected (F = 60.2). We observe similar results when we use *PercRevision* and *PropRevUp* as dependent variables. When we split *SignedExtreme* and *SignedModerate* intro their positive and negative components, we find that analysts react more strongly to extreme positive than to moderate positive language across all three specifications. However, extreme negative language has greater explanatory power than moderate negative language only in *PropRevUp* specification (F = 30.7). Taken together, these results suggest that analysts, similar to investors, appear to infer information from management's choice to use extreme rather than moderate language.

[29]Indeed, prior research documents that many, if not most, forecasts revisions happen either on the day of or the day after an earnings call (Clement et al. (2011), Huang et al. (2017)).

3.4 Extreme Language in Conference Calls and Future Operating Performance

Prior research on the role of language in corporate reporting suggests several ways in which managers might use extreme language. On one hand, managers may use extreme or moderate language to complement the underlying financial statements in hopes of better informing investors about current and future performance. On the other hand, managers may use extreme language for strategic purposes, such as masking poor performance or inflating moderately good performance. Alternatively, extreme language could largely reflect management or firm style and carry no special meaning beyond that. Prior research, in fact, finds some evidence for each of these roles. For example, the tone of management's qualitative disclosures is informative about future operating performance, both when looking at the MD&A section in annual and quarterly reports (Li (2010a), Bochkay and Levine (2017)) and when looking at earnings press releases (Davis et al., 2012). Huang et al. (2013) provide evidence that managers use qualitative statements to strategically manipulate investor opinion, and Davis et al. (2014) find that there is a significant management-specific component in earnings conference calls.

Our market reactions and analyst revisions results provide evidence that market participants pay attention to extreme language in earnings conference calls and, on average, find it informative. Therefore, our next test is to examine the extent to which there is support for such reactions when examining future operating performance. If linguistic extremity largely reflects management or firm style, we may find little or no information for predicting future operating performance. Alternatively, if managers use extreme positive language to hype their company or mislead investors, then we may see predictive power only in the moderate or negative language or find a negative association between extreme positive statements and future performance.[30] However, given the strong market and analyst reactions to linguistic extremity, we expect these reactions to be rooted, at least to some degree, in meaningful information about future fundamental performance. To test this prediction, we regress the

[30] The Private Securities Litigation Reform Act of 1995 provides strong safe harbor provisions for firms that provide meaningful cautionary disclaimers when making disclosures that contain forward-looking statements (Asay and Hales, 2018).

one-year-ahead change in earnings or sales relative to the current quarter, scaled by current total assets, on our signed variables for linguistic extremity:

$$\Delta Performance_{(q+4,q)} = \gamma_0 + \gamma_1 SignedExtreme + \gamma_2 SignedModerate + C \times Controls + \varepsilon, \quad (3)$$

where *Controls* include current earnings, growth in earnings over the last year, growth in sales over the last year, unexpected earnings, loss, loss interacted with the unexpected earnings, indicator variables for high and low unexpected earnings, pre-announcement return, accruals, earnings volatility, market-to-book, leverage, Altman's Z-score, stock returns, return volatility, analyst following, number of business and geographic segments, firm age, financial industry indicator, forward-looking disclosures in earnings calls, risk and uncertainty disclosures in earnings calls, and year-quarter fixed effects. If managers use linguistic extremity to inform rather than mislead investors about future performance, we expect to find a strong positive association with future earnings and sales ($\gamma_1 > 0$ and $\gamma_2 > 0$). Given our results in Tables V, VI and IX, we also expect *SignedExtreme* to be more informative than *SignedModerate* in predicting future firm performance ($\gamma_1 > \gamma_2$).

Table X reports the results. We find that both of types of language are significantly associated with one-year-ahead changes in sales, but only extreme language is associated with one-year-ahead changes in earnings. Moreover, a one standard deviation increase in *SignedExtreme* results in a 0.0020 (0.0026×0.777) increase in earnings scaled by assets and a 0.0037 (0.0026×1.449) increase in sales scaled by assets. A one standard deviation increase in *SignedModerate* increases sales by only 0.0016. These effects are large given that the median $\Delta Earn_{(q+4,q)}$ and $\Delta Sales_{(q+4,q)}$ in our sample is 0.0007 and 0.0075, respectively. The equality of the estimated coefficients on *SignedExtreme* and *SignedModerate* is strongly rejected for both earnings and sales (F = 39.1 and F = 28.2, respectively), suggesting that managers use extreme language to inform investors about future firm performance.

Next, we decompose *SignedExtreme* and *SignedModerate* into their positive and negative components. We find that extreme positive (negative) language is strongly positively (negatively) associated with future performance, while the association with moderate language is mostly insignificant. Further, the equality of the extreme and moderate coefficients is strongly rejected for both positive and negative language, suggesting greater informativeness of extreme language for future operating performance.[31] Overall, these results indicate that managers' use of extreme language contains information about future firm performance. In addition, the stronger market reactions to linguistic extremity, documented in Section 3.1, are, at least partially, warranted.

3.5 Information Content of Extremity vs. Tone in Earnings Conference Calls

In Tables V-X, we provide evidence on how markets react to linguistic extremity and what information it carries for future fundamental performance. We find that extreme words are more strongly associated with abnormal trading volume, stock returns, analyst forecast revisions, and future operating performance than moderate words, regardless of whether these words are classified as positive or negative. Even though Tables V-X do not have distinct variables to measure positive or negative tone of each word and its linguistic extremity, i.e., our measures of extremity are conditional on the initial measurement of tone, we interpret these results as the evidence that extremity of language matters as the split of tone into components is performed along the extremity dimension.

To provide further evidence that linguistic extremity captures an important, but a different attribute of language from tone, we provide a univariate evidence on the relationship between extreme language, tone, and abnormal returns. Specifically, Table XI summarizes the event and post-event period abnormal returns, *BHAR[0, 2]* and *BHAR[3, 60]*, for firms in *Low* and *High* unexpected earnings terciles, conditional on *Low* and *High* linguistic extremity, and *Low* and *High* linguistic tone. Unexpected earnings, extremity and tone terciles are created using quarterly independent sorts of earnings conference calls by the corresponding unexpected earnings (*UE*), extremity of the conference call (*SignedExtreme*), and tone of the conference call (*Tone*). Holding *Tone* fixed, we observe that market reactions to *UE* strongly depend on extremity of language in the conference call (*SignedExtreme*). For in-

[31] We get similar results when replacing our linguistic extremity measures with measures of abnormal extremity, which we calculate following a methodology analogous to what Huang et al. (2013) use for calculating abnormal tone. In addition, we find similar results if we include firm fixed effects and when we break the content of the whole conference call into the Prepared Remarks and Q&A sections.

stance, market reactions to low (high) *UE* is about 1.047% (2.274%) higher and significant at the 1% level when language is more extreme for the low level of *Tone* (Tone Tercile = 1). Similarly, market reactions to low (high) *UE* is about 2.140% (2.537%) higher and significant at the 1% level when language is more extreme for the high level of *Tone* (Tone Tercile = 3). These results are consistent with message tone (positive vs. negative) and message intensity (extreme vs. moderate) capturing complementary yet different information in earnings calls.

For post-announcement abnormal returns, we generally do not observe any significant return reversals following earnings conference calls. One exception is a 0.863% reversal (significant at the 10% level) for low unexpected earnings tercile when comparing firms with high vs. low extremity in the conference call. Figure 4 provides a graphical evidence on the relationship between event and post-event period abnormal returns, unexpected earnings, and extremity of language in earnings conference calls. Consistent with our results in Table XI, we observe higher market reactions to unexpected earnings when extremity is higher. Altogether, these univariate results provide additional evidence that extremity and tone capture different attributes of language in earnings conference calls and that investors price information in extreme language correctly.

While our results so far show that linguistic extremity is an important dimension of language, even when conditioned on whether those same words would be classified as positive or negative, they do not allow us to separately assess the importance of tone vs. extremity when analyzing earnings conference calls. Therefore, we next test the significance of extreme language, while *simultaneously* controlling for positive and negative tone. To have separate variables measuring tone and extremity, we first calculate traditional measures of positive and negative tone (i.e., number of positive words divided by total words, *Positive*; number of negative words divided by total words, *Negative*). Next, we measure the extent of extremity in positive (negative) language as the number of extreme positive (negative) words divided by total positive (negative) words, *ExtrWordsInPositive* (*ExtrWordsInNegative*). Measures of extremity constructed in this manner are not collinear with tone of the earnings call, but rather are more likely to capture the extent to which tonal words (positive and negative) are extreme.

Table XII reports the estimated coefficients from regressions of our main dependent variables (*CAV[0, 2]*, *BHAR[0, 2]*, *AmtOfRev*, *PercRev*, *PropRevUp*, $\Delta ROA[q+4,q]$ and $\Delta Sales[q+4,q]$) on positive and negative tone scores (*Positive* and *Negative*), extremity proportions within positive and negative words (*ExtrWordsInPositive* and *ExtrWordsInNegative*), and relevant control variables used in Tables V-X. We find that *ExtrWordsInPositive* and *ExtrWordsInNegative* have an incremental explanatory power in all of our market reactions, analyst revisions, and future performance models, as indicated by significant coefficient estimates in every specification. For instance, one standard deviation increase in *ExtrWordsInPositive* increases *BHAR[0, 2]* by 0.62%, while a standard deviation increase in *Positive* is associated with 0.15% higher *BHAR[0, 2]*. In addition, one standard deviation increase in *ExtrWordsInNegative* (*Negative*) results in 0.28% (0.65%) lower *BHAR[0, 2]*.

For each dependent variable in Table XII we also run an F-test to compare the goodness of fit of two models: (1) model that includes *Positive* and *Negative* tone in addition to *Controls* and (2) model that includes *ExtrWordsInPositive* and *ExtrWordsInNegative* measures in addition to *Positive*, *Negative* tone and *Controls*. In all instances, the F-test rejects the null of no incremental value of *ExtrWordsInPositive* and *ExtrWordsInNegative*. Overall, these results are consistent with our evidence in Tables V-X and suggest that market participants attend to tone and extremity as separate, but important, aspects of qualitative discussions in earnings conferences that accompany quantitative accounting information.

4 Conclusion

We use a large sample of earnings conference calls to develop a comprehensive dictionary for measuring extremity of spoken language. Each entry in our dictionary is ranked by human annotators according to its signed linguistic extremity (the extent to which it is positive/negative and extreme). We provide evidence that extreme language carries significant explanatory value for market reactions and future operating performance, above and beyond traditional measures of tone, performance and other firm characteristics. Specifically, we find that extreme language in earnings conference calls significantly increases trading volume and prompts strong price reactions in response to those calls. Both positive and negative extreme language results in significant market reactions. Further, market reactions to extreme language are more pronounced for firms with weaker information environments and higher information processing costs. We also find that analysts respond to extreme language as evidenced by their forecasting activities following earnings conference calls. When we examine an association between extreme language and future earnings and sales, we find that, on average, managers use extreme language to convey information about future operating performance. Moreover, investors seem to price information in extreme language correctly as we generally do not observe price reversals or drifts over the 60-day window following earnings conference calls. Taken together, our evidence suggests that market participants are influenced not just by what managers say, but also how they say it, with extreme language playing an important role in management communications with investors.

We believe our paper is the first large scale empirical study that shows that investors respond to extreme language in corporate disclosures. The extensive word coverage and word ratings in our dictionary open new avenues for future research. Researchers interested in conducting textual analysis in business settings (e.g., in IPO roadshows, social media contexts, the financial press, analyst reports, investor conferences, interviews, and more) could use our dictionary, particularly when spoken language is involved, to examine the extremity of language, its determinants, and its information content.

References

Andersen, P. A. and T. R. Blackburn (2004). An experimental study of language intensity and response rate in e-mail surveys. *Communication Reports 17*(2), 73.

Asay, H. S. and J. Hales (2018). Disclaiming the future: investigating the impact of cautionary disclaimers on investor judgments before and after experiencing economic loss. *The Accounting Review*.

Aune, R. K. and T. Kikuchi (1993). Effects of language intensity similarity on perceptions of credibility relational attributions, and persuasion. *Journal of Language and Social Psychology 12*(3).

Bamber, L. S., O. E. Barron, and D. E. Stevens (2011). Trading volume around earnings announcements and other financial reports: Theory, research design, empirical evidence, and directions for future research. *Contemporary Accounting Research 28*(2), 431–471.

Beaver, W. H. (1968). The information content of annual earnings announcements. *Journal of Accounting Research*, 67–92.

Bernard, V. L. and J. K. Thomas (1989). Post-earnings-announcement drift: delayed price response or risk premium? *Journal of Accounting Research*, 1–36.

Blankespoor, E. and E. DeHaan (2015). CEO visibility: Are media stars born or made? *Rock Center for Corporate Governance at Stanford University, Working Paper* (203).

Blankespoor, E., B. E. Hendricks, and G. S. Miller. Perceptions and price: Evidence from CEO presentations at IPO roadshows. *Journal of Accounting Research 55*(2), 275–327.

Blau, B. M., J. R. DeLisle, and S. M. Price (2015). Do sophisticated investors interpret earnings conference call tone differently than investors at large? Evidence from short sales. *Journal of Corporate Finance 31*, 203–219.

Bochkay, K. and C. B. Levine (2017). Using MD&A to improve earnings forecasts. *Journal of Accounting, Auditing & Finance*.

Bonsall, S. B. and B. P. Miller (2017). The impact of narrative disclosure readability on bond ratings and the cost of debt. *Review of Accounting Studies 22*(2), 608–643.

Bowers, J. W. (1963). Language intensity, social introversion, and attitude change. *Speech Monographs 30*(4), 345–352.

Brindley, B. and J. Law (2008). *A dictionary of Finance and Banking.* Oxford University Press.

Buhrmester, M., T. Kwang, and S. D. Gosling (2011). Amazon's mechanical turk a new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science 6*(1), 3–5.

Burgoon, M., S. B. Jones, and D. Stewart (1975). Toward a message-centered theory of persuation: three empirical investigations of language intensity. *Human Communication Research 1*(3), 240–256.

Campbell, C. J. and C. E. Wasley (1996). Measuring abnormal daily trading volume for samples of NYSE/ASE and NASDAQ securities using parametric and nonparametric test statistics. *Review of Quantitative Finance and Accounting 6*(3), 309–326.

Chambers, A. E. and S. H. Penman (1984). Timeliness of reporting and the stock price reaction to earnings announcements. *Journal of Accounting Research*, 21–47.

Chen, H., P. De, Y. J. Hu, and B.-H. Hwang (2014). Wisdom of crowds: the value of stock opinions transmitted through social media. *Review of Financial Studies 27*(5), 1367–1403.

Chen, S. and S. Chaiken (1999). The heuristic-systematic model in its broader context. *Dual-Process Theories in Social Psychology*, 73–96.

Chychyla, R., M. Minutti-Meza, and A. J. Leone (2018). Complexity of Financial Reporting Standards and Accounting Expertise. *Working Paper, University of Miami*.

Clement, M. B., J. Hales, and Y. Xue (2011). Understanding analysts' use of stock returns and other analysts' revisions when forecasting earnings. *Journal of Accounting and Economics 51*(3), 279–299.

Clementson, D. E., P. Pascual-Ferrá, and M. J. Beatty (2014). How language can influence political marketing strategy and a candidate's image: effect of presidential candidates' language intensity and experience on college students' ratings of source credibility. *Journal of Political Marketing*.

Daniel, K., M. Grinblatt, S. Titman, and R. Wermers (1997). Measuring mutual fund performance with characteristic-based benchmarks. *Journal of Finance*, 1035–1058.

Davis, A. K., W. Ge, D. Matsumoto, and J. L. Zhang (2014). The effect of manager-specific optimism on the tone of earnings conference calls. *Review of Accounting Studies 20*(2), 639–673.

Davis, A. K., J. M. Piger, and L. M. Sedor (2012). Beyond the numbers: measuring the information content of earnings press release language. *Contemporary Accounting Research 29*(3), 845–868.

Feldman, L., S. Govindaraj, J. Livnat, and B. Segal (2009). Managements tone change, post earnings announcement drift and accruals. *Review of Accounting Studies 15*(4), 915 – 953.

Frankel, R., M. Johnson, and D. J. Skinner (1999). An empirical examination of conference calls as a voluntary disclosure medium. *Journal of Accounting Research*, 133–150.

Frankel, R., W. J. Mayew, and Y. Sun (2010). Do pennies matter? Investor relations consequences of small negative earnings surprises. *Review of Accounting Studies 15*(1), 220–242.

Gilbert, D. T. (1991). How mental systems believe. *American Psychologist 46*(2), 107.

Gilbert, D. T., R. W. Tafarodi, and P. S. Malone (1993). You can't not believe everything you read. *Journal of Personality and Social Psychology 65*(2), 221.

Hales, J., X. J. Kuang, and S. Venkataraman (2011). Who believes the hype? An experimental examination of how language affects investor judgments. *Journal of Accounting Research 49*(1).

Hamilton, M. A. (1998). Message variables that mediate and moderate the effect of equivocal language on source credibility. *Journal of Language and Social Psychology 17*(1), 109–143.

Hanley, K. W. and G. Hoberg (2010). The information content of IPO prospectuses. *Review of Financial Studies 23*(7), 2821–2864.

Henry, E. and A. J. Leone (2015). Measuring qualitative information in capital markets research: Comparison of alternative methodologies to measure disclosure tone. *The Accounting Review 91*(1), 153–178.

Hirshleifer, D., S. S. Lim, and S. H. Teoh (2009). Driven to distraction: extraneous events and underreaction to earnings news. *The Journal of Finance 64*(5), 2289–2325.

Hoberg, G. and V. Maksimovic (2015). Redefining financial constraints: a text-based analysis. *Review of Financial Studies 28*(5), 1312–1352.

Huang, A., R. Lehavy, A. Zang, and R. Zheng (2017). Analyst information discovery and information interpretation roles: a topic modeling approach. *Management Science*.

Huang, X., S. H. Teoh, and Y. Zhang (2013). Tone management. *The Accounting Review 89*(3), 1083–1113.

Jegadeesh, N. and D. Wu (2013). Word power: a new approach for content analysis. *Journal of Financial Economics 110*(3), 712–729.

Jurafsky, D. and H. James (2000). Speech and language processing: an introduction to natural language processing, computational linguistics, and speech. *Pearson Education*.

Kimbrough, M. D. (2005). The effect of conference calls on analyst and market underreaction to earnings announcements. *The Accounting Review 80*(1), 189–219.

Kravet, T. and V. Muslu (2013). Textual risk disclosures and investors risk perceptions. *Review of Accounting Studies 18*(4), 1088–1122.

Larcker, D. F. and A. A. Zakolyukina (2012). Detecting deceptive discussions in conference calls. *Journal of Accounting Research 50*(2), 495–540.

Law, J. (2010). *A dictionary of accounting.* Oxford University Press.

Lee, J. (2015). Can investors detect managers' lack of spontaneity? *Adherence* to predetermined scripts during earnings conference calls. *The Accounting Review 91*(1), 229–250.

Li, F. (2008). Annual report readability, current earnings, and earnings persistence. *Journal of Accounting and Economics 45*(2 - 3), 221 – 247.

Li, F. (2010a). The information content of forward-looking statements in corporate filings - a naïve bayesian machine learning approach. *Journal of Accounting Research 48*(5), 1049–1102.

Li, F. (2010b). Textual analysis of corporate disclosures: a survey of the literature. *Journal of Accounting Literature 29*, 143–165.

Loughran, T. and B. McDonald (2011). When is a Liability not a Liability? *Journal of Finance 66*(1), 35 – 65.

Loughran, T. and B. McDonald (2016). Textual analysis in accounting and finance: a survey. *Journal of Accounting Research*.

Matsumoto, D., M. Pronk, and E. Roelofsen (2011). What makes conference calls useful? *The* information content of managers' presentations and analysts' discussion sessions. *The Accounting Review 86*(4), 1383–1414.

Milian, J. A. and A. L. Smith (2017). An investigation of analysts' praise of management during earnings conference calls. *Journal of Behavioral Finance 18*(1), 65–77.

Miller, B. P. (2010). The effects of reporting complexity on small and large investor trading. *The Accounting Review 85*(6), 2107–2143.

Monga, V. and E. Chasan (2015). CFO journal: the 109,894-word annual report - as regulators require more disclosures, 10-Ks reach epic lengths; how much is too much? *Wall Street Journal June 02*.

Muslu, V., S. Radhakrishnan, K. Subramanyam, and D. Lim (2014). Forward-looking MD&A disclosures and the information environment. *Management Science*.

Nisbett, R. E. and L. Ross (1980). Human inference: strategies and shortcomings of social judgment.

Osovsky, A. (2016). Puffery on the market: a behavioral economic analysis of the puffery defense in the securities arena. *Harvard Business Law Review*.

Paolacci, G., J. Chandler, and P. G. Ipeirotis (2010). Running experiments on amazon mechanical turk. *Judgment and Decision Making 5*(5), 411–419.

Petty, R. E. and D. T. Wegener (1999). The elaboration likelihood model: current status and controversies.

Price, S. M., J. S. Doran, D. R. Peterson, and B. A. Bliss (2012). Earnings conference calls and stock returns: the incremental informativeness of textual tone. *Journal of Banking & Finance 36*(4).

Rennekamp, K. (2012). Processing fluency and investors reactions to disclosure readability. *Journal of Accounting Research 50*(5), 1319–1354.

Rennekamp, K. M. and P. D. Witz (2017). Linguistic formality and perceived engagement–investors reactions to two unique characteristics of social media disclosures. *Working Paper*.

Rogers, J. L., A. V. Buskirk, and S. L. C. Zechman (2011). Disclosure tone and shareholder litigation. *The Accounting Review 86*(6), 2155–2183.

Taboada, M., J. Brooke, M. Tofiloski, K. Voll, and M. Stede (2011). Lexicon-based methods for sentiment analysis. *Computational Linguistics 37*(2), 267–307.

Tetlock, P. C. (2007). Giving Content to investor sentiment: the role of media in the stock Market. *The Journal of Finance 62*(3), 1139 – 1167.

Tetlock, P. C., M. Saar-Tsechansky, and S. Macskassy (2008). More than words: quantifying language to measure firms' fundamentals. *The Journal of Finance 63*(3), 1437 – 1467.

You, H. and X. Zhang (2009). Financial reporting complexity and investor underreaction to 10-K information. *Review of Accounting Studies 14*(4), 559 – 586.

Tables

Table I: Examples of Words and Phrases in the Rated Dictionary.

Rating	Words and Phrases	Extracts from Conference Calls
5	excellence, superior, terrific, wonderful, top quality, exceptional, best, incredible, extremely well, amazing	we continued to produce exceptional results; we are able to achieve superior outcomes because of our ability to share best practices; we experienced phenomenal growth rates; we do have some amazing, amazing innovation; this is terrific for us and our shareholders especially because the incremental value starts benefiting us this quarter.
4	strong, success, tremendous, really good, exceed expectations, powerful, leader, very good, great, prosper	our operating cash flow remains strong; we have experienced continuing success; we are both a powerful marketing vehicle as well as a commerce vehicle; we feel very good about the job we are doing; I remain confident in our company's ability to grow and prosper.
3	accomplishment, solid, improvement, work hard, effective, strengthen, optimistic, healthy, proud, high quality	we are proud of this accomplishment; we have had solid business across our businesses; we managed to strengthen our financial structure; we are optimistic about the future; we have a healthy balance sheet.
2	increase, growth, please, be able to, gain, expand, move forward, improve, continue to deliver, advantage	we are pleased with our operating and financial performance; this increase is due to revenue growth; we continue to improve the operating margins; we continue to expand our capacity; we do have an advantage over most other companies.
1	generate, competitive, in line, produce, lower cost, steady, encourage, transparency, sufficient, a bit better	our view would be to generate capital through sales; the performance is in line with our expectation; we have seen steady volume; we have a sufficient cash generation; we are going to be as competitive as we need to be.
-1	issue, force, limitation, expensive, complexity, heavy, step back, undue, unexpected, not on	the issue is the pricing of products; this effort will be expensive; revenues deferred due to project complexity; we experienced unexpected changes in revenues; it as a limitation on what we could do.
-2	weak, slow, slowdown, delay, concern, decrease, uncertain, adversely, work against, go down	we got off to a slow start; we expect a seasonal slowdown in volume; the tone of business has been a particular concern; we remain uncertain about the demand; our operating results were adversely affected by.
-3	loss, difficult, volatile, underperform, diminish, hard, fall short, unfavorable, decline, be behind	our consumer business has just completed a difficult season; the quarter was certainly more volatile than normal; we reported an operating loss of; we expect our volumes to underperform; we are having a hard time catching up.
-4	failing, weakness, negative, suffer, disappoint, deteriorate, disruptive, sharp decline, unsuccessful, get worse	we experienced continued weakness in our business; we still suffer declines in our international business; we have, to date, been unsuccessful; we are going to be disruptive in the market; the negative impact was more than anticipated.
-5	default, terrible, horrible, worst, devastate, bankrupt, very bad, very poor, in serious trouble, out of business	our results this year were the worst; we are close to being bankrupt; this quarter's performance has been horrible; we have had terrible spud to sales; it takes an awful long time to get projects underway.

Table II: Variable Definitions and Data Sources.

Variable	Definition	Source
PosExtreme	Proportion of extreme positive words used in the earnings conference call (see Section 2).	Earnings calls are from: www.seekingalpha.com
PosModerate	Proportion of moderate positive words used in the earnings conference call (see Section 2).	Earnings calls are from: www.seekingalpha.com
NegModerate	Proportion of moderate negative words used in the earnings conference call (see Section 2).	Earnings calls are from: www.seekingalpha.com
NegExtreme	Proportion of extreme negative words used in the earnings conference call (see Section 2).	Earnings calls are from: www.seekingalpha.com
TotalExtreme	Sum of *PosExtreme* and *NegExtreme*.	
TotalModerate	Sum of *PosModerate* and *NegModerate*.	
SignedExtreme	Difference between *PosExtreme* and *NegExtreme*.	
SignedModerate	Difference between *PosModerate* and *NegModerate*.	
Positive	Proportion of positive words used in the earnings conference call.	Earnings calls are from: www.seekingalpha.com
Negative	Proportion of negative words used in the earnings conference call.	Earnings calls are from: www.seekingalpha.com
Tone	Difference between *Positive* and *Negative*.	
ExtrWordsInPositive	Proportion of extreme positive words to total positive words in the earnings conference call (see Section 2).	Earnings calls are from: www.seekingalpha.com
ExtrWordsInNegative	Proportion of extreme negative words to total negative words in the earnings conference call (see Section 2).	Earnings calls are from: www.seekingalpha.com
$BHAR[0,2]$	Cumulative abnormal return over the 3-day event window calculated relative to a size, book-to-market, and momentum matching portfolio return (Daniel et al., 1997), where day 0 is the conference call date.	CRSP
$CAV[0,2]$	Log-transformed cumulative abnormal volume over the 3-day event window calculated relative to the expectation as in Campbell and Wasley (1996), where day 0 is the conference call date.	CRSP
$\Delta ROA[q+4,q]$	Earnings before extraordinary items in quarter $q+4$ minus earnings before extraordinary items in the current quarter q divided by total assets in the current quarter q, winsorized at 1% and 99%.	COMPUSTAT
$\Delta Sales[q+4,q]$	Sales in quarter $q+4$ minus sales in the current quarter q divided by total assets in the current quarter q, winsorized at 1% and 99%.	COMPUSTAT
AmountOfRevision	Average analysts' forecast revision, where forecast revision for an individual analyst is calculated as his first quarterly estimate of four quarters ahead EPS issued within the 10-day window following the earnings call, less his previous estimate, scaled by the median stock price of the month preceding the announcement, winsorized at 1% and 99%.	IBES
PercRevision	Average relative analysts' forecast revision, where forecast revision for an individual analyst is calculated as his first quarterly estimate of four quarters ahead EPS issued within the 10-day window following the earnings call, less his previous estimate, scaled by the previous estimate, winsorized at 1% and 99%.	IBES

Table II, continued

Variable	Definition	Source
PropRevUp	Proportion of analysts revising their forecasts upwards within the 10-day window following the earnings call, winsorized at 1% and 99%.	IBES
UE	Actual earnings per share (EPS) minus analyst consensus forecast of one- or two-quarters-ahead earnings issued or reviewed in the last 60 days before earnings announcement divided by stock price at the end of quarter, winsorized at 1% and 99%.	IBES
High UE	Indicator variable that equals to 1 if *UE* is in the highest decile in a given quarter.	
Low UE	Indicator variable that equals to 1 if *UE* is in the lowest decile in a given quarter.	
Loss	Indicator variable that equals to 1 if actual EPS is lower than 0.	IBES
PreAnnRet	Cumulative pre-announcement return calculated using daily returns between the analyst forecast date and two days before earnings announcement.	CRSP
ROA	Earnings before extraordinary items scaled by total assets, winsorized at 1% and 99%.	COMPUSTAT
Accruals	Earnings minus cash flows from operations divided by book value of assets, winsorized at 1% and 99%.	COMPUSTAT
Size	Natural logarithm of the market value of equity at the end of the previous quarter.	COMPUSTAT
MTB	Market value of equity plus book value of liabilities divided by book value of assets measured at the end of the previous quarter, winsorized at 1% and 99%.	COMPUSTAT
Leverage	Long-term debt to total assets ratio.	COMPUSTAT
ZScore	Altman's Z-Score.	COMPUSTAT
EarnVol	Standard deviation of earnings, calculated using earnings scaled by total assets in the last twenty quarters, with a minimum of eight quarters required.	COMPUSTAT
RetVol	Standard deviation of monthly returns, calculated using returns in the last twelve month, with a minimum of six months required.	CRSP
NumAnalysts	Natural logarithm of the number of analysts that issue an earnings forecast for a given firm.	IBES
BusGeoSeg	Nuatural logarithm of the number of business and geographic segments.	COMPUSTAT
FirmAge	Natural logarithm of the number of years since a company appears in the CRSP's monthly file.	CRSP
FLS	Proportion of sentences in the earnings conference call containing a forward-looking term from Muslu et al. (2014).	Earnings calls are from: www.seekingalpha.com
Risk	Proportion of sentences in the earnings conference call containing at least one of the risk related terms from Kravet and Muslu (2013).	Earnings calls are from: www.seekingalpha.com
Uncertainty	Proportion of uncertain words in the earnings conference call. List of uncertain words is from Loughran and McDonald (2011).	Earnings calls are from: www.seekingalpha.com

Table III: Descriptive Statistics.

	Mean	Median	SD	Q1	Q3
Main Independent Variables					
PosExtreme	0.0074	0.0072	0.0025	0.0056	0.0090
PosModerate	0.1599	0.1598	0.0142	0.1500	0.1697
NegModerate	0.0289	0.0285	0.0054	0.0249	0.0324
NegExtreme	0.0010	0.0008	0.0007	0.0005	0.0013
TotalExtreme	0.0085	0.0082	0.0025	0.0066	0.0101
TotalModerate	0.1889	0.1890	0.0129	0.1802	0.1978
SignedExtreme	0.0064	0.0062	0.0026	0.0045	0.0081
SignedModerate	0.1308	0.1309	0.0172	0.1188	0.1430
Positive	0.1672	0.1672	0.0165	0.1567	0.1780
Negative	0.0300	0.0295	0.0060	0.0258	0.0336
ExtrWordsInPositive	0.0442	0.0431	0.0133	0.0344	0.0529
ExtrWordsInNegative	0.0322	0.0284	0.0212	0.0166	0.0436
Main Dependent Variables					
CAV[0,2]	1.6114	1.5263	1.6325	0.5680	2.5657
BHAR[0,2]	0.0002	−0.0008	0.0906	−0.0433	0.0435
AmtOfRevision	−0.0007	0.0001	0.0059	−0.0013	0.0011
PercRevision	−0.0080	0.0083	0.3269	−0.0763	0.0781
PropRevUp	0.5177	0.5000	0.3871	0.1111	1.0000
$\Delta ROA[q+4,q]$	−0.0001	0.0007	0.0444	−0.0065	0.0072
$\Delta Sales[q+4,q]$	0.0124	0.0075	0.0591	−0.0096	0.0311
Controls					
UE	0.0001	0.0005	0.0117	−0.0006	0.0022
Loss	0.1692	0.0000	0.3750	0.0000	0.0000
AbsBHAR	0.0626	0.0433	0.0656	0.0192	0.0840
ROA	0.0036	0.0114	0.0486	0.0008	0.0221
Accruals	−0.0175	−0.0140	0.0425	−0.0297	−0.0001
Size	7.6049	7.5881	1.7473	6.3461	8.7620
MTB	2.0902	1.6237	1.4576	1.2118	2.3942
Leverage	0.1902	0.1674	0.1750	0.0133	0.2977
ZScore	3.5751	2.2255	5.6181	1.1526	4.1212
EarnVol	0.0284	0.0139	0.0385	0.0070	0.0328
RetVol	0.1128	0.0976	0.0635	0.0677	0.1403
NumAnalysts	1.9907	1.9459	0.7572	1.3863	2.5649
BusGeoSeg	1.6283	1.6094	0.5766	1.0986	2.0794
FirmAge	2.4643	2.5870	1.1069	1.7545	3.3474
FLS	0.1251	0.1205	0.0420	0.0948	0.1502
Risk	0.1040	0.1007	0.0288	0.0839	0.1206
Uncertainty	0.0096	0.0094	0.0023	0.0080	0.0109

This table shows the descriptive statistics for variables used in the paper over the period 2006-2015. All variables are defined in Table II.

Table IV: Correlation Table

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
(1) *TotalExtreme*	1.00													
(2) *TotalModerate*	0.20***	1.00												
(3) *SignedExtreme*	0.82***	0.22***	1.00											
(4) *SignedModerate*	0.33***	0.78***	0.45***	1.00										
(5) *Positive*	0.41***	0.89***	0.49***	0.94***	1.00									
(6) *Negative*	−0.25***	−0.03***	−0.48***	−0.63***	−0.40***	1.00								
(7) *PropExtrInPos*	0.91***	−0.04***	0.90***	0.16***	0.21***	−0.31***	1.00							
(8) *PropExtrInNeg*	0.28***	−0.06***	−0.26***	−0.09***	−0.08***	0.19***	0.03***	1.00						
(9) *CAV[0,2]*	0.05***	0.02***	0.06***	0.06***	0.05***	−0.07***	0.05***	−0.01	1.00					
(10) *BHAR[0,2]*	0.09***	0.00	0.13***	0.06***	0.05***	−0.10***	0.11***	−0.06***	−0.04***	1.00				
(11) *AmtOfRevision*	0.12***	0.01*	0.16***	0.10***	0.08***	−0.14***	0.14***	−0.04***	0.02***	0.25***	1.00			
(12) *PercRevision*	0.07***	−0.01**	0.10***	0.05***	0.03***	−0.09***	0.09***	−0.03***	−0.01	0.15***	0.23***	1.00		
(13) *PropRevUp*	0.17***	0.03***	0.23***	0.14***	0.12***	−0.20***	0.19***	−0.07***	−0.03***	0.32***	0.55***	0.30***	1.00	
(14) $\Delta ROA[q+4,q]$	0.01	0.03***	0.03***	0.02***	0.03**	−0.07***	0.02***	−0.02***	−0.05***	−0.01	−0.02***	−0.01	0.03***	1.00
(15) $\Delta Sales[q+4,q]$	0.05***	0.02	0.09***	0.08***	0.05***	−0.13***	0.06***	−0.06***	−0.01	0.07***	0.10***	0.05***	0.11***	0.23***

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

This table shows Pearson correlations between our main variables of interest. All variables are defined in Table II.

Table V: Extreme Language in Earnings Conference Calls and Event Period Cumulative Abnormal Volume (CAV[0, 2]).

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
TotalExtreme	38.183***		36.287***		30.004***		28.171***	
	(5.45)		(5.07)		(6.25)		(5.89)	
TotalModerate		3.435***	2.183**			3.108***	2.204**	
		(3.16)	(1.98)			(3.05)	(2.20)	
PosExtreme				31.249***				26.850***
				(3.83)				(5.61)
PosModerate				2.170*				2.213**
				(1.91)				(2.09)
NegModerate				−11.543***				−3.577
				(−3.53)				(−1.23)
NegExtreme				24.911				34.704**
				(1.32)				(2.09)
Controls	No	No	No	No	Yes	Yes	Yes	Yes
F-test of								
Extreme = Moderate			20.67^{+++}				27.35^{+++}	
Positive				11.32^{+++}				22.33^{+++}
Negative				3.29^{+}				4.95^{++}
Observations	35,155	35,155	35,155	35,155	35,155	35,155	35,155	35,155
Adj. R^2	0.055	0.052	0.055	0.057	0.258	0.256	0.258	0.258

This table shows the estimated coefficients from a regression of the 3-day event period cumulative abnormal volume on extreme and moderate language scores. *Controls* include: *Announcement News* = [*UE*, *Loss*, *Loss* × *UE*, *High UE*, *Low UE*, *AbsBHAR*]; *Earnings* = [*ROA*, *Accruals*, *EarnVol*]; *Firm Characteristics* = [*Size*, *MTB*, *Leverage*, *ZScore*, *RetVol*, *NumAnalysts*, *BusGeoSeg*, *FirmAge*, *FinInd*]; *Conf. Call Content* = [*FLS*, *Risk*, *Uncertainty*]. Year-quarter fixed effects and the constant are included in each regression, but are not reported. All variables are defined in Table II. ***, **, * indicate significance at the 1%, 5%, and 10% levels, respectively, using two-tailed t-test. Reported statistics are based on a two-way clustering at both firm level and year-quarter level. $^{+++}$, $^{++}$, $^{+}$ indicate significance at the 1%, 5%, and 10% levels, respectively, using an F-test.

Table VI: Extreme Language in Earnings Conference Calls and Event Period Abnormal Returns (BHAR[0, 2]).

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
SignedExtreme	4.433***		4.203***		3.694***		3.337***	
	(18.91)		(17.80)		(17.19)		(16.53)	
SignedModerate		0.362***	0.087**			0.348***	0.147***	
		(10.53)	(2.43)			(9.51)	(4.09)	
PosExtreme				3.473***				2.802***
				(14.67)				(13.70)
PosModerate				−0.132***				−0.054
				(−3.10)				(−1.25)
NegModerate				−1.016***				−1.074***
				(−7.18)				(−10.07)
NegExtreme				−8.014***				−5.330***
				(−11.41)				(−7.85)
Controls	No	No	No	No	Yes	Yes	Yes	Yes
F-test of								
Extreme = *Moderate*			268.67^{+++}				235.29^{+++}	
Positive				219.52^{+++}				190.15^{+++}
Negative				97.96^{+}				37.30^{++}
Observations	35,155	35,155	35,155	35,155	35,155	35,155	35,155	35,155
Adj. R^2	0.017	0.005	0.018	0.022	0.095	0.089	0.096	0.099

This table shows the estimated coefficients from a regression of the 3-day event period buy-and-hold abnormal return on extreme and moderate language scores. *Controls* include: *Announcement News* = [*UE*, *Loss*, *Loss* × *UE*, *High UE*, *Low UE*]; *Earnings* = [*ROA*, *Accruals*, *EarnVol*]; *Firm Characteristics* = [*Size*, *MTB*, *Leverage*, *ZScore*, *PreAnnRet*, *RetVol*, *NumAnalysts*, *BusGeoSeg*, *FirmAge*, *FinInd*]; *Conf. Call Content* = [*FLS*, *Risk*, *Uncertainty*]. Year-quarter fixed effects and the constant are included in each regression, but are not reported. All variables are defined in Table II. ***, **, * indicate significance at the 1%, 5%, and 10% levels, respectively, using a two-tailed t-test. Reported statistics are based on a two-way clustering at both firm level and year-quarter level. $^{+++}$, $^{++}$, $^{+}$ indicate significance at the 1%, 5%, and 10% levels, respectively, using an F-test.

Table VII: Extreme Language in Conference Calls and Event Period Cumulative Abnormal Volume (CAV[0, 2]), sorts by firm characteristics.

	Size		*NumAnalysts*		*NumInstOwn*	
TotalExtreme	25.420***	22.068***	27.631***	23.344***	19.656***	18.771***
	(3.66)	(4.29)	(3.67)	(4.29)	(2.91)	(3.68)
TotalModerate	1.869	2.851***	2.428*	2.123*	3.488***	2.691***
	(1.48)	(2.69)	(1.87)	(1.83)	(2.92)	(2.69)
Group	−0.599	−0.329	−0.156	−0.306	−0.281	−0.474
	(−1.43)	(−0.89)	(−0.45)	(−0.93)	(−0.68)	(−1.37)
TotalExtreme × *Group*	43.456***	26.448**	24.919**	14.820*	33.923***	21.685**
	(3.34)	(2.28)	(2.09)	(1.89)	(2.84)	(2.41)
TotalModerate× *Group*	1.399	−1.235	−0.303	0.517	−1.137	0.505
	(0.62)	(−0.66)	(−0.16)	(0.31)	(−0.52)	(0.28)
Controls	No	Yes	No	Yes	No	Yes
Observations	34,939	34,939	34,939	34,939	34,939	34,939
Adj. R^2	0.056	0.263	0.055	0.259	0.063	0.265

This table shows the estimated coefficients from a regression of the 3-day event period buy-and-hold abnormal volume on extreme and moderate language scores for small and large, low and high analyst following, and low and high institutional ownership firms. Each quarter firms are sorted into three groups based on their size (small, medium, and large), analyst following (low, medium, and high), and the number of institutional owners (low, medium, and high). *Group* equals 1 if a firm belongs to a small, low analyst following, and low institutional ownership group in the first, second, and third columns, respectively, and 0 otherwise. *Controls* is a vector of control variables used in Table V. Year-quarter fixed effects, and the constant are included in the regressions, but are not reported. All variables are defined in Table II. ***, **, * indicate significance at the 1%, 5%, and 10% levels, respectively, using the two-tailed t-test. Reported statistics are based on a two-way clustering at both firm level and year-quarter level.

Table VIII: Extreme Language in Conference Calls and Event Period Abnormal Returns (BHAR[0, 2]), sorts by firm characteristics.

	Size		*NumAnalysts*		*NumInstOwn*	
SignedExtreme	3.382***	2.852***	3.566***	2.929***	3.131***	2.635***
	(12.66)	(11.23)	(12.47)	(10.45)	(12.26)	(10.98)
SignedModerate	0.124***	0.180***	0.104**	0.176***	0.135***	0.206***
	(3.30)	(5.66)	(2.20)	(4.28)	(3.27)	(5.92)
Group	−0.002	0.008	−0.006	0.006	−0.014	−0.001
	(−0.17)	(0.85)	(−0.63)	(0.72)	(−1.60)	(−0.15)
SignedExtreme × *Group*	2.993***	1.612***	1.747***	0.995**	3.063***	2.078***
	(5.15)	(3.14)	(3.39)	(2.03)	(5.39)	(4.22)
SignedModerate × *Group*	−0.114	−0.110	−0.044	−0.084	−0.073	−0.118*
	(−1.20)	(−1.26)	(−0.53)	(−1.09)	(−1.07)	(−1.85)
Controls	No	Yes	No	Yes	No	Yes
Observations	34,939	34,939	34,939	34,939	34,939	34,939
Adj. R^2	0.019	0.096	0.018	0.095	0.019	0.099

This table shows the estimated coefficients from a regression of the 3-day event period buy-and-hold abnormal return on extreme and moderate language for small and large, low and high analyst following, and low and high institutional ownership firms. Each quarter firms are sorted into three groups based on their size (small, medium, and large), analyst following (low, medium, and high), and number of institutional owners (low, medium, and high). *Group* equals 1 if a firm belongs to a small, low analyst following, and low institutional ownership group in the first, second, and third columns, respectively, and 0 otherwise. *Controls* is a vector of control variables used in Table VI. Year-quarter fixed effects, and the constant are included in the regressions, but are not reported. All variables are defined in Table II. ***, **, * indicate significance at the 1%, 5%, and 10% levels, respectively, using the two-tailed t-test. Reported statistics are based on a two-way clustering at both firm level and year-quarter level.

Table IX: Extreme Language in Conference Calls and Subsequent Analyst Forecast Revisions.

	AmtOfRev				*PercRev*				*PropRevUp*			
SignedExtreme	0.300***		0.175***		10.934***		7.992***		28.702***		23.036***	
	(13.81)		(8.80)		(11.80)		(8.30)		(15.67)		(13.63)	
SignedModerate	0.006		0.008**		−0.018		0.073		1.120***		1.428***	
	(1.40)		(2.44)		(−0.10)		(0.47)		(4.77)		(6.23)	
PosExtreme		0.279***		0.166***		9.700***		7.013***		25.202***		19.841***
		(13.08)		(9.46)		(10.20)		(7.38)		(13.36)		(11.61)
PosModerate		−0.008*		−0.001		−0.718***		−0.519***		−0.260		0.187
		(−1.75)		(−0.16)		(−3.60)		(−2.61)		(−1.02)		(0.79)
NegModerate		−0.071***		−0.053***		−3.092***		−2.754***		−7.607***		−7.574***
		(−4.06)		(−3.56)		(−5.47)		(−4.84)		(−13.66)		(−14.33)
NegExtreme		−0.273***		−0.142**		−11.742***		−9.139**		−37.046***		−31.087***
		(−3.92)		(−2.21)		(−2.70)		(−2.08)		(−8.07)		(−7.42)
Controls	No	No	Yes	Yes	No	No	Yes	Yes	No	No	Yes	Yes
F-test of												
Extreme= Moderate	158.9^{+++}		60.2^{+++}		118.1^{+++}		57.1^{+++}		196.6^{+++}		143.0^{+++}	
Positive		161.3^{+++}		77.1^{+++}		105.4^{+++}		54.4^{+++}		160.9^{+++}		116.6^{+++}
Negative		6.9^{+++}		1.5		3.7^{+}		1.9		40.9^{+++}		30.7^{+++}
Observations	24,338	24,338	24,338	24,338	24,338	24,338	24,338	24,338	24,338	24,338	24,338	24,338
Adj. R^2	0.045	0.048	0.168	0.170	0.018	0.020	0.048	0.050	0.080	0.087	0.140	0.146

This table shows the estimated coefficients from a regression of analyst forecast revisions and probability of forecast upgrade following the conference call on extreme and moderate language scores. First, second, and last four columns report coefficient estimates for the amount of forecast revision scaled by price (*AmtOfRev*), forecast revision in percent (*PercRev*), and the proportion of upward forecast revisions (*PropRevUp*), respectively. *Controls* include: *Announcement News* = [*UE*, *Loss*, *Loss* × *UE*, *High UE*, *Low UE*]; *Earnings* = [*ROA*, *Accruals*, *EarnVol*]; *Firm Characteristics* = [*Size*, *MTB*, *Leverage*, *ZScore*, *PreAnnRet*, *RetVol*, *NumAnalysts*, *BusGeoSeg*, *FirmAge*, *FinInd*]; *Conf. Call Content* = [*FLS*, *Risk*, *Uncertainty*]. Year-quarter fixed effects and the constant are included in each regression, but are not reported. All variables are defined in Table II. ***, **, * indicate significance at the 1%, 5%, and 10% levels, respectively, using the two-tailed t-test. Reported statistics are based on a two-way clustering at both firm level and year-quarter level. $^{+++}$, $^{++}$, $^{+}$ indicate significance at the 1%, 5%, and 10% levels, respectively, using an F-test.

Table X: Extreme Language in Conference Calls and Future Performance.

	$\Delta ROA[q+4,q]$				$\Delta Sales[q+4,q]$			
SignedExtreme	0.036		0.777***		1.141***		1.449***	
	(0.35)		(6.90)		(4.19)		(6.28)	
SignedModerate	0.001		0.002		0.158**		0.096*	
	(0.05)		(0.08)		(2.47)		(1.76)	
PosExtreme		0.161*		0.718***		0.542*		1.099***
		(1.72)		(6.66)		(1.79)		(4.48)
PosModerate		0.069**		0.018		−0.059		−0.007
		(2.43)		(0.59)		(−1.13)		(−0.17)
NegModerate		0.304***		0.082		−1.173***		−0.571***
		(3.63)		(1.20)		(−5.74)		(−2.90)
NegExtreme		0.094		−1.503***		−2.498***		−3.040***
		(0.19)		(−2.95)		(−3.07)		(−4.33)
Controls	No	No	Yes	Yes	No	No	Yes	Yes
F-test of								
Extreme = Moderate	0.12		39.1^{+++}		9.78^{+++}		28.2^{+++}	
Positive		1.08		37.7^{+++}		3.30^{+}		17.8^{+++}
Negative		0.16		8.65^{+++}		2.05		10.1^{+++}
Observations	36,171	36,171	36,171	36,171	36,171	36,171	36,171	36,171
Adj. R^2	0.034	0.035	0.284	0.284	0.092	0.100	0.148	0.150

This table shows the estimated coefficients from a regression of one-year-ahead change in earnings scaled by total assets ($\Delta ROA[q+4,q]$) and one-year-ahead change in sales scaled by total assets ($\Delta Sales[q+4,q]$) on extreme and moderate language scores. *Controls* include: *Announcement News* = [*UE*, *Loss*, *Loss* × *UE*, *High UE*, *Low UE*]; *Earnings* = [*ROA*, $\Delta ROA[q,q-4]$, $\Delta Sales[q,q-4]$, *Accruals*, *EarnVol*]; *Firm Characteristics* = [*Size*, *MTB*, *Leverage*, *ZScore*, *PreAnnRet*, *RetVol*, *NumAnalysts*, *BusGeoSeg*, *FirmAge*, *FinInd*]; *Conf. Call Content* = [*FLS*, *Risk*, *Uncertainty*]. All variables are defined in Table II. Year-quarter fixed effects and the constant are included in each regression, but are not reported. ***, **, * indicate significance at the 1%, 5%, and 10% levels, respectively, using the two-tailed t-test. Reported statistics are based on a two-way clustering at both firm level and year-quarter level. $^{+++}$, $^{++}$, $^{+}$ indicate significance at the 1%, 5%, and 10% levels, respectively, using an F-test.

Table XI: Buy-and-Hold Abnormal Returns, Sorted by Unexpected Earnings, Linguistic Extremity and Tone of Earnings Conference Calls

	BHAR[0, 2]			BHAR[3, 60]		
UE \ **Extremity**	**Low**	**High**	**High-Low**	**Low**	**High**	**High-Low**
Panel A. Overall						
Low	-3.824% [7,166]	-2.372% [4,063]	1.451%*** (7.42)	-0.215% [7,166]	0.647% [4,063]	0.863%* (1.84)
High	2.007% [4,488]	4.241% [5,350]	2.233%*** (9.17)	-0.120% [4,488]	0.811% [5,350]	0.932% (1.50)
High-Low	5.831%*** (21.2)	6.614%*** (19.1)		0.095% (0.18)	0.164% (0.26)	
Panel B. Tone Tercile = 1						
Low	-3.654% [4,369]	-2.607% [596]	1.047%*** (2.68)	-0.183% [4,369]	2.212% [596]	2.395%* (1.78)
High	1.870% [2,483]	4.144% [701]	2.274%*** (5.93)	-0.195% [2,483]	0.498% [701]	0.715% (0.75)
High-Low	5.524%*** (19.2)	6.751%*** (18.4)		-0.034% (-0.05)	-1.714% (-0.95)	
Panel C. Tone Tercile = 2						
Low	-3.980% [2,036]	-2.471% [1,385]	1.509%*** (5.52)	0.460% [2,036]	0.652% [1,385]	0.192% (0.23)
High	2.231% [1,408]	3.651% [1,652]	1.419%*** (3.96)	-0.022% [1,408]	0.533% [1,652]	-0.555% (0.68)
High-Low	6.212%*** (13.8)	6.122%*** (14.7)		-0.482% (-0.56)	-0.119% (-0.18)	
Panel D. Tone Tercile = 3						
Low	-4.380% [761]	-2.240% [2,082]	2.140%*** (4.95)	-2.227% [761]	0.194% [2,082]	2.422%** (2.59)
High	2.052% [597]	4.589% [2,997]	2.537%*** (4.74)	0.045% [597]	1.039% [2,997]	0.993% (0.85)
High-Low	6.432%*** (9.93)	6.830%*** (16.1)		2.272%* (1.67)	0.844% (1.51)	

This table shows the average 3-day event period buy-and-hold abnormal return (*BHAR[0, 2]*) and 58-day post-event period buy-and-hold abnormal return (*BHAR[3, 60]*) for high and low earnings surprise terciles (*Low UE*: bad news; *High UE*: good news) by high and low linguistic extremity terciles (high and low *SignedExtreme*). Panel A reports results unconditional on *Tone* of the earnings call, while Panels B through D report results for low (tercile 1), medium (tercile 2), and high (tercile 3) *Tone* of the earnings call. Day 0 is the date of the earnings conference call. Earnings surprise, tone and extremity terciles are created using quarterly independent sorts of earnings conference calls by the corresponding unexpected earnings (*UE*), tone (*Tone*) and extremity of the conference call (*SignedExtreme*). All measures are defined in Table II. T-statistics based on a two-way clustering at both firm level and year-quarter level (number of observations) are in parenthesis (squared brackets).

Table XII: Extreme Language in Conference Calls, Controlling for Tone.

	Market Reactions		Analyst Revisions			Future Performance	
	CAV[0, 2]	*BHAR[0, 2]*	*AmtOfRev*	*PercRev*	*PropRevUp*	$\Delta ROA[q+4,q]$	$\Delta Sales[q+4,q]$
Positive	2.848***	0.091**	0.006**	−0.060	1.077***	0.046*	0.044
	(3.33)	(2.29)	(2.01)	(−0.38)	(5.93)	(1.85)	(1.35)
Negative	−2.938	−1.092***	−0.052***	−2.609***	−7.737***	0.024	−0.612***
	(−1.10)	(−10.98)	(−4.10)	(−5.13)	(−15.71)	(0.41)	(−3.46)
ExtrWordsInPositive	4.390***	0.473***	0.028***	1.195***	3.221***	0.117***	0.179***
	(5.09)	(14.18)	(8.96)	(7.01)	(10.63)	(6.41)	(4.13)
ExtrWordsInNegative	1.039*	−0.130***	−0.010***	−0.392**	−0.718***	−0.045***	−0.071***
	(1.96)	(−5.55)	(−3.01)	(−2.16)	(−5.55)	(−2.85)	(−2.95)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
F-test of Incremental Effects of Extremity Scores	29.7+++	86.9+++	45.4+++	23.9+++	149.4+++	31.9+++	37.4+++
Observations	35,155	35,155	24,338	24,338	24,338	36,171	36,171
Adj. R^2	0.258	0.099	0.170	0.050	0.146	0.283	0.150

This table shows the estimated coefficients from a regression of main dependent variables of interest (*CAV[0, 2]*, *BHAR[0, 2]*, *AmtOfRev*, *PercRev*, *PropRevUp*, $\Delta ROA[q+4,q]$ and $\Delta Sales[q+4,q]$) on positive and negative tone scores (*Positive* and *Negative*), extremity proportions within positive and negative words (*ExtrWordsInPositive* and *ExtrWordsInNegative*), and relevant control variables used in Tables V-X. All variables are defined in Table II. Year-quarter fixed effects and the constant are included in each regression, but are not reported. ***, **, * indicate significance at the 1%, 5%, and 10% levels, respectively, using the two-tailed t-test. Reported statistics are based on a two-way clustering at both firm level and year-quarter level. +++, ++, + indicate significance at the 1%, 5%, and 10% levels, respectively, using an F-test for the incremental explanatory power of *ExtrWordsInPositive* and *ExtrWordsInNegative* relative to benchmark models that exclude these variables.

Figures

Figure 1: Illustration of a Human Intelligence Task (HIT) Designed to Rate Dictionary Words and Phrases.

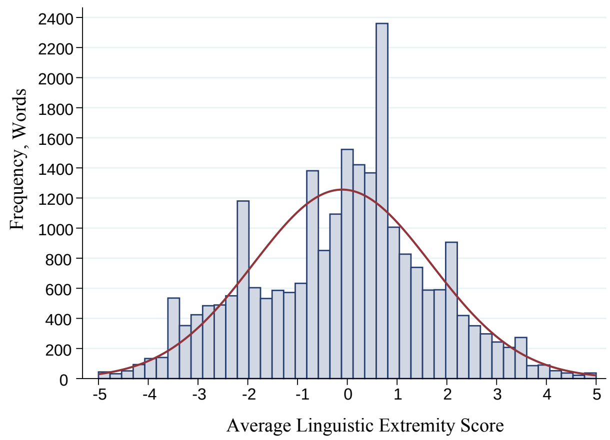
In the context of firm management describing recent company performance, how positive or negative are the following words and phrases?

Word/Phrase	Unable To Rate	Strongly Negative (-5)	(-4)	Negative (-3)	(-2)	(-1)	Neutral (0)	(1)	(2)	Positive (3)	(4)	Strongly Positive (5)
phenomenal												
forgery												
be proud of												
deleterious												
satisfy												
distressed												
evolve												
strengthening												
poor quality												
assist												
important												
decline												

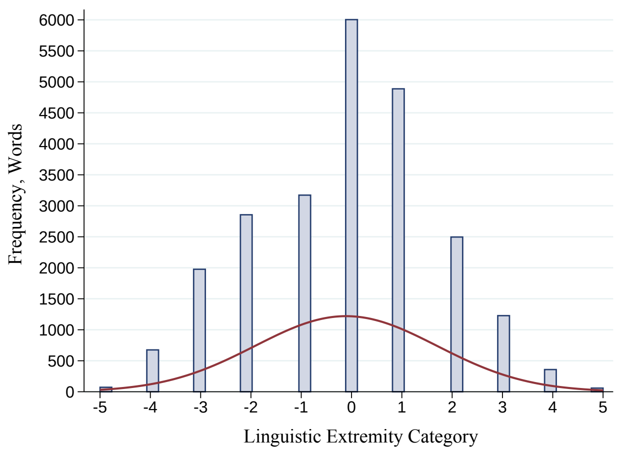
This figure presents a snippet of a Human Intelligence Task (HIT) developed through Amazon Mechanical Turk service to rate a set of words and phrases to identify their signed linguistic extremity. Each HIT is designed to have human annotators rate 50 unique randomly selected words and phrases on a scale from (-5) to (5), where (-5) is "extremely negative", (0) is "neutral", and (5) is "extremely positive". Raters are allowed to select "Unable to Rate" if they are not familiar with the word and are unable to provide a meaningful rating. Each HIT is rated by 5 different raters. To increase the quality of responses, each HIT contains several 'attention check' questions.

Figure 2: Frequency Distribution of Words based on Linguistic Extremity Ratings.

(a)

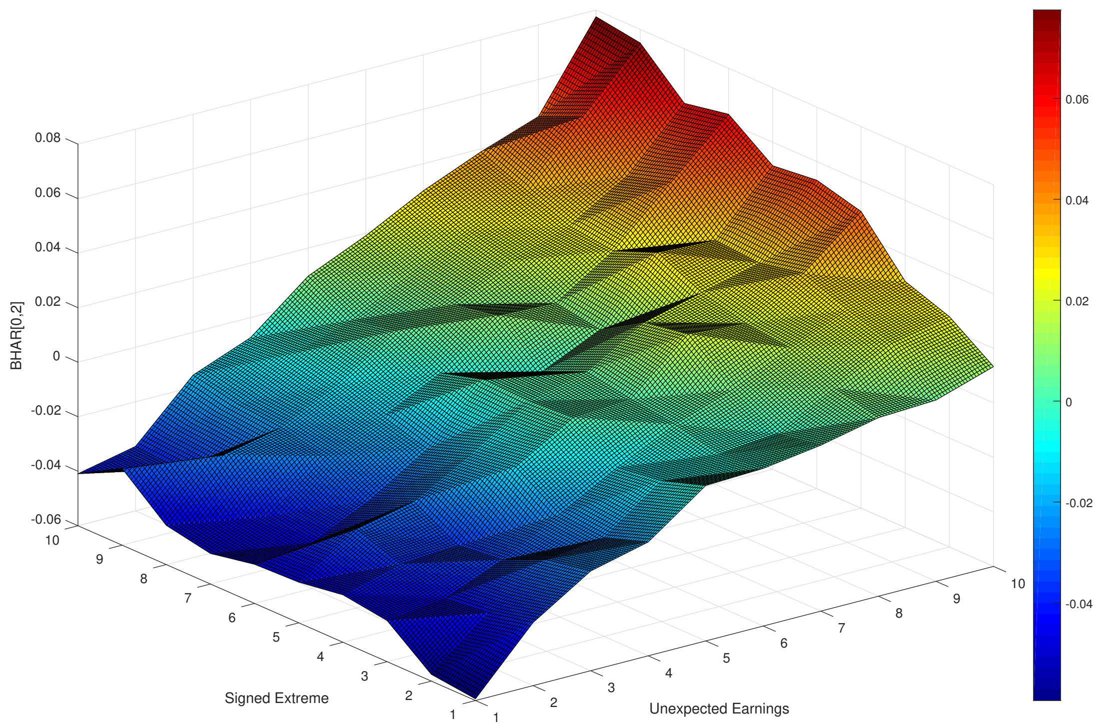


(b)



This figure shows the frequency distribution of linguistic extremity ratings of 23,355 words and phrases in *DICT*. Each rating is the average of five individual ratings collected by employing human annotators through Amazon Mechanical Turk service. Raters were asked to indicate how negative or positive each word or phrase is on a scale from (-5) for "extremely negative" to (+5) for "extremely positive", where (0) indicates a neutral word. Figure (a) plots the frequency distribution of a continuous linguistics extremity measure of words and phrases in *DICT*, whereas Figure (b) plots the frequency distribution of a categorical linguistic extremity measure, where each average rating was rounded.

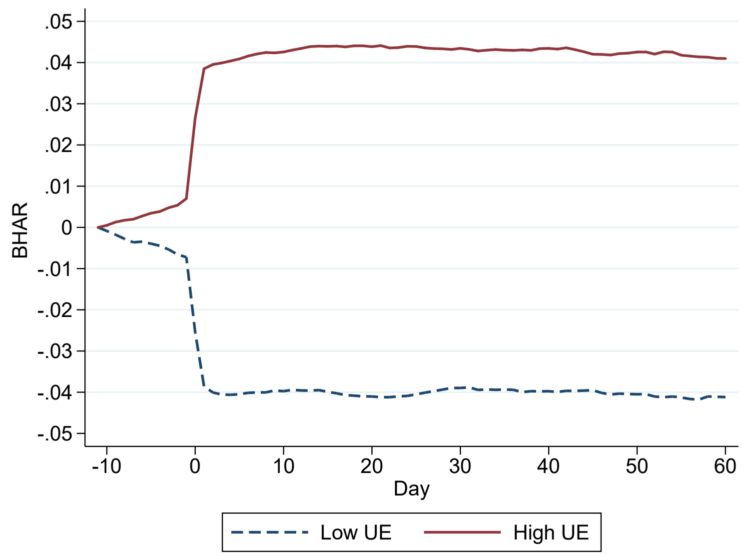
Figure 3: Event Period Buy-and-Hold Abnormal Returns and Extreme Tone of Earnings Conference Calls.



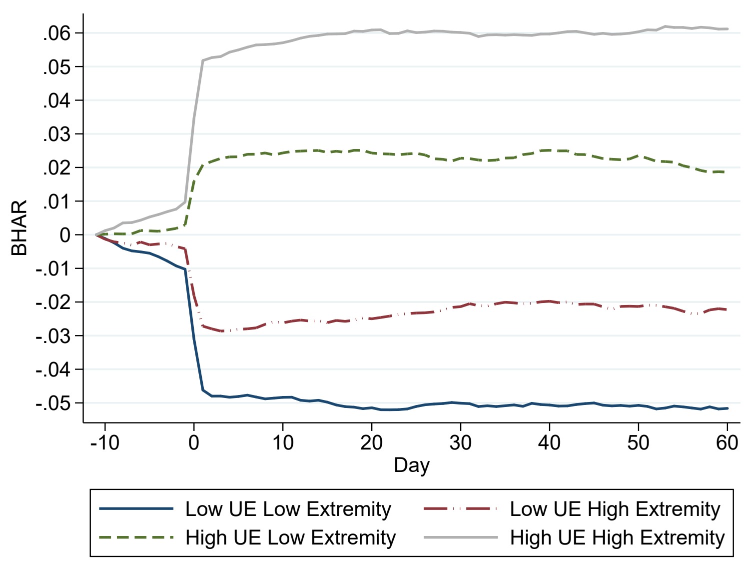
This figure shows the event-period cumulative buy-and-hold abnormal return (BHAR[0, 2]) of quarterly earnings announcements against earnings surprise (*UE*) deciles (1: bad news, 10: good news) and deciles of language extremity (*SignedExtreme*) in the accompanying conference call (1: low extremity; 10: high extremity). Earnings surprise and extremity deciles are created using quarterly independent double sorts of *UE* and *SignedExtreme*. All measures are defined in Table II.

Figure 4: Abnormal Returns of High and Low Earnings Surprise Firms by High and Low Extreme Tone of Earnings Conference Calls

(a) Sort by Unexpected Earnings



(b) Sort by Unexpected Earnings and Earnings Call Extremity



This figure plots cumulative buy-and-hold abnormal returns (*BHAR*) following quarterly earnings announcements accompanied by earnings conference calls. Panel (a) plots *BHAR* for low and high earnings surprise terciles (*Low UE*: bad news; *High UE*: good news). In Panel (b), observations in *Low UE* and *High UE* are further grouped by low and high *SignedExtreme* terciles. Earnings surprise and extremity terciles are created using quarterly independent double sorts of quarterly earnings calls by the corresponding unexpected earnings (*UE*) and signed extremity of the conference call (*SignedExtreme*). All measures are defined in Table II. The time period starts 10 days prior to the conference call and ends 60 days after the call. Day 0 is the date of the earnings conference call.

Internet Appendix

Table IA.1: Extreme Language in Earnings Conference Calls and Event Period Cumulative Abnormal Volume (CAV[0, 2]), Introductory Remarks and Q&A Sections.

	Intro	*QA*	*Intro&QA*
TotalExtremeIntro	16.632***		12.376***
	(5.46)		(4.12)
TotalModerateIntro	3.324***		3.176***
	(5.39)		(5.21)
TotalExtremeQA		25.512***	17.942***
		(5.78)	(4.17)
TotalModerateQA		2.290**	0.454
		(2.40)	(0.49)
Controls	Yes	Yes	Yes
F-test of *Extreme = Moderate*			
Intro	17.6+++		8.58+++
Q&A		26.2+++	15.5+++
Observations	35,155	35,155	35,155
Adj. R^2	0.259	0.257	0.259

This table shows the estimated coefficients from a regression of the 3-day event period cumulative abnormal volume on extreme and moderate language scores in the introductory remarks and Q&A sections of the conference call. Control variables include: *Announcement News* = [*UE*, *Loss*, *Loss* × *UE*, *High UE*, *Low UE*, *AbsBHAR*]; *Earnings* = [*ROA*, *Accruals*, *EarnVol*]; *Firm Characteristics* = [*Size*, *MTB*, *Leverage*, *ZScore*, *RetVol*, *NumAnalysts*, *BusGeoSeg*, *FirmAge*, *FinInd*]; *Conf. Call Content* = [*FLS*, *Risk*, *Uncertainty*]. Year and quarter fixed effects and the constant are included in each regression, but are not reported. All variables are defined in Table II. ***, **, * indicate significance at the 1%, 5%, and 10% levels, respectively, using the two-tailed t-test. Reported statistics are based on a two-way clustering at both firm level and year-quarter level. +++, ++, + indicate significance at the 1%, 5%, and 10% levels, respectively, using an F-test.

Table IA.2: Extreme Language in Earnings Conference Calls and Event Period Abnormal Returns (BHAR[0, 2]), Introductory Remarks and Q&A Sections.

	Intro	*QA*	*Intro&QA*
SignedExtremeIntro	1.680***		1.123***
	(13.31)		(9.33)
SignedModerateIntro	0.053***		−0.028
	(2.67)		(−1.30)
SignedExtremeQA		2.954***	2.397***
		(15.56)	(12.96)
SignedModerateQA		0.225***	0.214***
		(6.09)	(5.26)
Controls	Yes	Yes	Yes
F-test of *Extreme = Moderate*			
Intro	160.2+++		87.7+++
Q&A		197.4+++	133.7+++
Observations	35,155	35,155	35,155
Adj. R^2	0.091	0.095	0.096

This table shows the estimated coefficients from a regression of the 3-day event period buy-and-hold abnormal return on extreme and moderate language scores in the introductory remarks and Q&A sections of the conference call. Control variables include: *Announcement News* = [*UE*, *Loss*, *Loss* × *UE*, *High UE*, *Low UE*]; *Earnings* = [*ROA*, *Accruals*, *EarnVol*]; *Firm Characteristics* = [*Size*, *MTB*, *Leverage*, *ZScore*, *PreAnnRet*, *RetVol*, *NumAnalysts*, *BusGeoSeg*, *FirmAge*, *FinInd*]; *Conf. Call Content* = [*FLS*, *Risk*, *Uncertainty*]. Year and quarter fixed effects and the constant are included in each regression, but are not reported. All variables are defined in Table II. ***, **, * indicate significance at the 1%, 5%, and 10% levels, respectively, using the two-tailed t-test. Reported statistics are based on a two-way clustering at both firm level and year-quarter level. +++, ++, + indicate significance at the 1%, 5%, and 10% levels, respectively, using an F-test.