LISTENING IN ON INVESTORS' CONVERSATIONS

Hailiang Chen and Byoung-Hyoun Hwang*

This Draft: January 2020

Utilizing server log data from one of the most popular investment-related websites in the United States, we study what types of stock-opinion articles investors, themselves, find the most captivating and more frequently read to the very end and what types of articles investors end up sharing with their peers through email. We find that the types of articles with the highest number of read-to-ends and the types of articles with the highest number of shares often reflect diametrically opposed spectra in terms of content attributes. For instance, while investors prefer reading ideas of more negative overall sentiment, it is the ideas of more positive sentiment that they share more frequently. Results from additional analyses suggest that such "sharing preferences" have implications for both investors' trading performances and asset prices.

JEL Classification: G11, G12, G14, G20.

Keywords: Investor Communication, Social Transmission Bias, Asset Prices.

---

* Chen is affiliated with the Faculty of Business and Economics, The University of Hong Kong, Hong Kong. Hwang is affiliated with the Cornell SC Johnson College of Business, Dyson School of Applied Economics and Management, Cornell University, Warren Hall 310E, Ithaca, NY 14853, USA. E-mail: chen19@hku.hk and bhwang@cornell.edu. The authors would like to thank Bing Han, Soo Kim, Wei Jiang, Justin Murfin and seminar participants at Baruch College, Cornell University, Korea University and the University of Hong Kong for many valuable suggestions and comments.

1. Introduction

Behavioral finance has become an oft-proposed alternative to the traditional framework in finance (Hirshleifer 2001; Barberis and Thaler 2003). In essence, behavioral finance asks whether the behavior of investors and financial markets is better described using models that are "psychologically more realistic" compared with those relied upon in the traditional framework.

At this point, it appears safe to conclude that behavioral finance has been helpful in explaining basic facts about the trading behavior of investors, the cross-section of average stock returns, and the aggregate stock market (Barberis and Thaler 2003). Behavioral finance has also shown that it can make concrete, out-of-sample predictions, some of which have already been confirmed in the data (Barberis, 2018).

Despite the many successes that behavioral finance can claim, some have advocated "a need to move from behavioral finance to *social finance*" (Hirshleifer 2015, p. 151; Hirshleifer 2020). The argument goes as follows: The manner in which behavioral models are made psychologically more realistic is by accounting for some of the many "behavioral biases" that individuals are prone to exhibit. The major biases that behavioral finance focuses on, such as extrapolative beliefs or prospect-theory preferences, all represent biases at the individual-person level. Investors do not operate in a vacuum, however, and it appears plausible that biases exist not only at the individual-person level, but also in the manner in which individuals communicate with one another. In particular, individuals may distort information. They may also not share all they know and feel and, instead, communicate only select types of content. To signal competence, for instance, investors may prefer to share stories of investment successes over stories of investment failures. To the degree that "receivers" do not discount for such communication decision and treat what they hear as representative of the "senders'" information sets, individuals end up misperceiving reality and potentially making suboptimal decisions. In our particular example, receivers may erroneously infer that outperforming the market is easy and flock to active investment strategies.

A large body of work finds evidence that investors, indeed, do not operate in a vacuum.[1] Lerner and Malmendier (2013), Shue (2013) and Fracassi (2017) find evidence that word of mouth exists also among corporate managers. Less is known about the *select types of content* that investors prefer sharing with each other and which of these "sharing preferences" are the potentially most useful in explaining the behavior of investors and financial markets.[2] This paper reports on a preliminary attempt to improve on this situation.

The primary challenge facing empirical work on word of mouth is that we generally do not observe investors' interactions with each other. Such data constraint has become less binding with the advent of modern information technologies. As our information gathering and our conversations shift online, for better or worse, we can increasingly track individuals' activities through the digital footprints they leave behind.

In our particular case, we obtain server log data from Seeking Alpha (hereafter, SA; http://seekingalpha.com). SA is one of the biggest investment-related websites in the United States. Users can submit stock-opinion articles to SA for possible publication. If deemed of adequate quality and published on the SA website, articles can generate income for the corresponding authors through the number of page views such articles generate.[3] For the period running from August 2012 through March 2013 and for every article published on the SA website, we have data on how often a reader scrolled to the bottom of an article ("number of read-to-ends") and how often an article was shared via email ("number of shares").

In the first part of our analyses, we use these data to describe what types of content investors prefer sharing. Owing to the richness of our data, we can also shed light on whether the types of content that

[1] For instance, Shiller and Pound (1989), Hong, Kubik and Stein (2004), Brown, Ivkovic, Smith and Weisbenner (2008), Kaustia and Knuepfer (2012), Banerjee, Chandrasekhar, Duflo and Jackson (2013), Hvide and Ostberg (2015), Pool, Stoffman and Yonker (2015), Heimer (2016).

[2] Kaustia and Knupfer (2012), Heimer and Simon (2017) and Huang, Hwang and Lou (2019) provide evidence that investors are more likely to share their investment experience when they have done well in the past, and Bali, Hirshleifer, Peng and Tang (2019) find evidence that retail investors enjoy conversing about lottery-like stocks.

[3] Some articles also receive cash awards. During our sample period, contributors were paid merely by the number of pages views. These days, contributor compensation is a joint function of page views and "reader scores." More details regarding how contributors are compensated can be found here: https://seekingalpha.com/pages/article_payments.

investors prefer sharing also are the types of content that investors, themselves, find the most captivating and more frequently read to end.

Loosely motivated by word-of-mouth studies in social psychology and marketing (surveyed in Lovett, Peres, and Shachar (2013), Berger (2014, 2016)), we consider the following attributes of a stock opinion article to gauge what types of content investors prefer sharing: sentiment, extremeness of viewpoint, uniqueness, emotionality, quantitative nature and usefulness. We capture all these attributes through textual analysis of the articles and data on how SA editors internally rated each SA article. We also consider whether a given investment idea pertains to a stock that resides in the "long legs" or the "short legs" of anomalies.[4]

The perhaps most surprising result from the first part of our analysis is that the types of stories that investors, themselves, find the most captivating and more frequently read to end and the types of stories they end up sharing with the "outside world" often reflect diametrically opposed spectra of content attributes. In particular, while investors more frequently read ideas to the end if they are of more negative overall sentiment, greater emotionality and lower reliance on numbers, it is the more positive, impersonal and quantitative ideas that generate the higher numbers of shares and thus become a greater part of investors' conversations. Similarly, while investors enjoy reading about stocks residing in the long leg of anomalies, they have little interest in conversing about such stocks with fellow investors.

A long line of work rooted in social psychology and marketing examines why individuals engage in word of mouth and why some stories, products and brands become contagious while others fade away (Berger 2014). A key thesis in this literature is that individuals do not necessarily share content that they, personally, find the most interesting. Instead, they prefer disseminating information that makes them feel better and signals thoughtfulness, knowledge and intelligence. Consistent with such "emotion regulation-" and "impression management" desires being important to investors, Kaustia and Knupfer (2012), Heimer

[4] Short- and long-leg stocks are generally stocks that reside at either extreme of a given firm characteristic. Prior work finds evidence that short-leg stocks subsequently significantly underperform long-leg stocks (e.g. Engelberg, McLean, and Pontiff 2018; Chen and Zimmermann 2019). One possible explanation is that the former tend to be overpriced while the latter tend to be underpriced. Such mispricing subsequently corrects, explaining the observed subsequent performance differential (e.g. Hirshleifer 2001; Barberis and Thaler 2003; Engelberg, McLean, and Pontiff; Chen and Zimmermann).

and Simon (2017) and Huang, Hwang and Lou (2019) provide evidence that investors are much more likely to share their investment experience when they have done well in the past than when they have done poorly. Our observed discrepancy between "content consumed" and "content shared" thus does not come as a complete surprise: Investors may be reluctant to share negative content because they do not want to be perceived as a negative person (Tesser and Rosen 1975; Berger and Milkman 2012; Forest and Wood 2012). Similarly, investors' preference for sharing impersonal, quantitative content may reflect efforts to appear balanced, detail-oriented, thoughtful and intelligent (Berger 2014).

The second part of our study considers possible implications for both investors' trading performances and asset prices. First, we examine whether the above sharing preferences have the potential to hurt investors' trading performances and, if so, which of the above sharing preferences are the potentially most harmful. We detail our analysis and results in the main body of the text. In short, when contrasting the types of articles that investors more frequently read to end to the types of articles that investors share more frequently, our results suggest that while the former strongly help earn abnormal returns, the latter do not. Simply put, investors keep the more useful content to themselves. We find hints in the data that the sharing preferences most responsible for this performance wedge include (1) investors' reluctance to share content that is mostly qualitative and more emotional and (2) investors' prepossession (disinterest) in discussing stocks residing in the short leg (long leg) of anomalies.

To examine possible asset pricing implications, we draw from the anomalies literature and construct *Short Score*, which is the number of anomalies in which a stock resides in the short leg, and *Long Score*, which is the number of anomalies in which a stock resides in the long leg. Consistent with findings reported in prior literature (e.g. Engelberg, McLean, and Pontiff 2018; Chen and Zimmermann 2019), we find that stocks with high *Short Score* (*Long Score*) earn low (high) returns going forward.

A common interpretation of this finding is that stocks with high *Short Score* carry characteristics that overly excite some investors. This causes temporary overpricing among short-leg securities. The more exciting features a stock possesses, the more overpriced it becomes. Overpricing subsequently corrects and produces low returns going forward (e.g. Hirshleifer 2001; Barberis and Thaler 2003; Engelberg, McLean and Pontiff 2018; Chen and Zimmermann 2019). The analogue argument applies to stocks with high *Long Score*.

In this paper, we speculate that a stock's exhibiting exciting features is not in itself a sufficient condition for the stock to become overpriced. To generate overpricing, one needs the presence of both exciting features and the sharing of those exciting features by investors.

Our results strongly corroborate this speculation. *Short Score* positively associates with contemporaneous returns and negatively predicts returns only among the subset of stocks with frequent sharing. Among stocks that attract infrequent sharing, a higher *Short Score* does not come with higher contemporaneous returns; neither does it precede lower subsequent returns, no matter how high the *Short Score*. These results suggest that investor communication along with investors' preferences for certain types of topics are relevant forces in the stock market.

Our study contributes to at least two strands of the literature. Our first and primary contribution is to the finance literature and the literature on social finance, in particular. The two building blocks of social finance are (1) that investors derive much of their information through word of mouth and (2) that investors communicate only select types of content and listeners do not fully account for such communication decisions; this causes investors to misperceive reality, make suboptimal investment decisions and distort asset prices.[5]

While there is strong evidence for the first building block, we know relatively little about what select types of content investors hear (or do not hear) from each other in the first place. Here, we provide a description of such investor sharing preferences, most of which, we believe, are new to the literature. We also provide preliminary evidence that these sharing preferences are important in explaining investor behavior and asset prices. Overall, we believe our paper may serve as useful reference for future work gauging to what degree social finance helps explain basic patterns in investor behavior and financial markets.

[5] As alluded to earlier, investors may also distort information.

Our paper should also be of interest to audiences outside of finance. The paper most closely related to ours is that of Berger and Milkman (2012). Berger and Milkman consider whether more positive, more emotional and more useful *New York Times* articles are more likely to make the *New York Times'* most emailed list. As we discuss in the main body of the text, some of the findings of Berger and Milkman carry over to our investor setting; others do not and we provide possible explanations. More importantly, unlike Berger and Milkman, we have server log data and can observe not only whether an article made a "most emailed list," but exactly how many times an article was actually shared. We can also observe how many times an article was read to end. Owing to such information richness, we believe our paper represents the first systematic documentation that the stories that capture our own attention and the stories that we end up sharing with the outside world often reflect diametrically opposed spectra of content attributes. We also provide evidence that the types of content we prefer keeping to ourselves are more useful than the types of content that we end up sharing.

2. Data and Variables

2.1. Data

The backbone of our analysis comprises server log data provided by SA. SA is a leading investments-related website in the United States. Any individual can submit a stock opinion article for possible publication on the SA website. These submissions are curated by a team of SA editors. SA suggests that as of August 2019 there have been 16,600 contributors and close to 1 million articles.[6] The SA website attracts over 15 million unique visitors a month, who, on average, spend seven minutes per visit.[7] The SA website thus has heavier web traffic than popular sites such as cnbc.com (11.4 million) and weather.com (11 million).[8]

[6] https://seekingalpha.com/listing/contributors_stats and https://seekingalpha.com/listing/articles_stats
[7] https://seekingalpha.com/page/who_reads_sa
[8] https://www.quantcast.com/top-sites/

Our server log data cover the period from August 1, 2012 through March 31, 2013. For each *article ID*, which is an identifier set by SA to uniquely identify an SA article, we have data on how often a reader scrolled to the bottom of such article and how often such article was shared via email.[9] Our server log data also contain scores that SA editors assigned internally to each article based on how "actionable," "convincing," or "well-presented" they perceived an article to be (described further in Section 2.2).

We augment our server log data with additional article-level data, which we scrape directly from the SA website. The information we scrape for each article includes the *article ID* and title, the full article text, the date of publication, the author's name, and the stock ticker. Some SA articles focus on one stock while others discuss multiple stocks. In this study, we conjecture that the virality of an investment idea is tied to characteristics of the focal stock. We therefore focus our analysis on single-stock opinion articles. Our final sample comprises 21,483 single-stock opinion articles, of which 16,446 contain the data necessary to construct all the variables used in our primary regression equation.

2.2 Variables

Dependent Variables

We have two dependent variables. Our first variable measures the degree to which an investment idea grabs investors' attention. In particular, we compute the natural logarithm of one plus the number of times an article is read to end, *ln (1 + # Read-to-Ends)*. Our second variable measures the level of virality of an investment idea and equals the natural logarithm of one plus the number of times an article is shared via email, *ln (1 + # Shares)*. In some regression specifications, we consider *ln (1 + # Shares)* conditional on an article having been read to end; in other specifications, we consider *ln (1 + # Shares)* by itself. We take the natural logarithm because *# Read-to-Ends* and *# Shares* are highly right-skewed.

[9] Our data are as of March 2014. Since 88.3% of article-reads in our sample occur in the first week of article publication (or 95.9% in the first month of article publication), any observed difference in article-reads and shares should not represent differences in time since article publication.

Independent Variables

The development of our independent variables is geared towards explaining the level of virality of an investment idea, *# Shares*, as, from our perspective, this is our primary dependent variable. Our independent variables are loosely motivated within the "theoretical framework" of Berger (2014). Related frameworks can be found in Berger (2016) and Lovett, Peres, and Shachar (2013).

A key point in the above frameworks is that senders do not engage in word of mouth with the sole purpose of assisting their receivers. A primary motivation for individuals to engage in word of mouth is to make themselves look or feel better. In particular, Berger (2014) argues that word of mouth serves the following five functions: (1) Impression management: Individuals share information to shape the impressions others have of them in a manner that agrees with each individual's "desired self." For instance, an individual may share a story about a secret bar to signal that she is hip, unique, and interesting. (2) Emotion regulation: Individuals use word of mouth to regulate their emotions. For instance, sharing positive stories about a recent vacation can help individuals savor past positive emotional experiences. On the other end of the spectrum, venting about a negative experience may provide some emotional catharsis. (3) Social bonding: Word of mouth can help individuals connect with others, help them feel part of a larger group and reduce loneliness. (4) Persuasion: Individuals use word of mouth to persuade others. For instance, an individual may share a story about a friend's positive family vacation experience to persuade her family to take a similar family vacation. (5) Information acquisition: Finally, individuals still do engage in word of mouth to provide and seek advice.

Berger (2014) builds on the five abovementioned functions to explain the types of content that individuals prefer (or avoid) sharing and the environmental conditions that make word-of-mouth effects stronger. We focus our discussion here on the determinants we deem most applicable to the contagion of ideas in financial markets.

a) Sentiment: Tesser and Rosen (1975) and Berger and Milkman (2012) suggest that people prefer sharing positive content over negative content as they do not want to be perceived as a negative or pessimistic person. In line with this view, Berger and Milkman find evidence that more positive *New York Times* articles are more likely to make the *New York Times'* most emailed list. Relatedly, using Facebook data, Forest and Wood (2012) find evidence that posting negative content can lead people to be liked less. Applying these insights to our particular setting, we may expect investors to prefer sharing stories about "buy opportunities" to stories about imminent market corrections.[10]

Somewhat counter to this prediction is evidence that individuals are viewed as more thoughtful and competent if they provide more critical advice (Amabile 1983). Moreover, Dubois, Bonezzi and DeAngelis (2016) suggest that the sharing of information with an individual close to the sender, as is presumably the case when a SA user forwards an article through email, "*activates a motive to protect others (Cross, Bacon, and Morris 2000; Cross and Madson 1997; Heine et al. 1999)*" and that "*sharing negative information is typically instrumental to consumers' motive to protect others (Hennig-Thurau et al. 2004; Sundaram, Mitra, and Webster 1998)*" (page 713). In the end, whether, ex ante, investors prefer sharing ideas of the more positive type or of the more negative type is not entirely clear.

To assess the effect of sentiment on sharing, we follow Berger and Milkman (2012) and construct *Sentiment* as the number of positive words in a SA article minus the number of negative words, scaled by the total number of words. As has become the norm for textual analysis in the field of finance, we use the lists of positive and negative words of Loughran and McDonald (2011).

b) Extremeness of Viewpoint: Whatever the direction of the effect of sentiment on sharing, we suspect that it is unlikely to be linear. Instead, we suspect kinks at the extremes as disseminating extreme views, irrespective of whether such views are positive or negative, may make the sharer appear either as more interesting and unique or as less balanced and thoughtful.

[10] Most retail investors buy stocks and do not engage in short-selling (Barber and Odean 2007). A stock market crash thus represents a threat to most retail investors and not an investment opportunity.

To assess whether extreme viewpoints are shared more (or less) frequently, we construct *Extreme Sentiment*, which equals one if *Sentiment* is either in the top or bottom decile of its distribution, and zero otherwise.

c) Uniqueness: We suspect that unique content is shared more frequently as disseminating such content makes the sharer appear more interesting and knowledgeable.

For each article written on stock *j* by author *k*, we construct both the average *Sentiment* across all articles written on stock *j* over the previous month, $Sentiment_{Stock}$, and the average *Sentiment* across all articles composed by author *k* over the previous month, $Sentiment_{Author}$. Our measures of uniqueness, $Unusual\ Sentiment_{Stock}$ and $Unusual\ Sentiment_{Author}$, are the absolute differences between the sentiment of the article in question and its corresponding $Sentiment_{Stock}$ and $Sentiment_{Author}$, respectively.

d) Emotionality: Emotion regulation is considered one of the key determinants of sharing and a large body of work argues and finds evidence that emotional content is shared more frequently than non-emotional content (Luminet, Bouts, Delie, Manstead, and Rime 2000; Peters, Kashima, and Clark 2009; Berger and Milkman 2012). High emotionality also triggers emotional arousal, which further encourages sharing (Dichter 1966).

Other research suggests that emotional advice is viewed as less balanced, less thoughtful, and less believable than non-emotional advice (Vendemia 2017). This may discourage the sharing of emotional content for impression-management purposes. In the end, similar to sentiment and extremeness of viewpoint, we deem it plausible that emotionality is a strong determinant of virality. But whether, in the aggregate, investors prefer sharing emotional content or impersonal content is an empirical question.

To assess the effect of emotionality on sharing, we follow Berger and Milkman (2012) and construct *Emotionality* as the number of positive words in a SA article plus the number of negative words scaled by the total number of words. We again use the lists of negative and positive words of Loughran and McDonald (2011).

e) Quantitative Nature: Opinions couched in numbers are often viewed as more thoughtful and more convincing (Koetsenruiter 2011). Statements backed up by numbers may also be seen as more precise. We thus suspect that investors prefer sharing opinions of more quantitative nature.

We measure the reliance on numbers versus text, *Numbers-versus-Text*, as the ratio of the total occurrences of numbers in a SA article to the total number of words.

f) Usefulness: We speculate that content perceived to be more useful is shared more frequently as sharing such content makes the sharer appear more thoughtful and helpful (Berger and Milkman 2012). Naturally, such content is also of greater interest to those seeking advice.

We measure usefulness through scores assigned by SA editors. Each SA article in our sample is rated by SA editors based on (1) how convincing, (2) how actionable and (3) how well-presented they perceive the article to be. Scores range from 1 through 5 with 1 being the lowest and 5 being the highest. SA's explanation of their scores is as follows: "*<u>Convincing</u> means that the writer of the article understands the topic he/she is writing about and the writer has subject matter expertise. It shows the writer can share pertinent information about the stock, sector, or style of investing. <u>Actionable</u> means that the writer of the article informed the investors and provided new information about a security or sector to help empower the investors with a better perspective on whether or not they should take a position. The value of the information the writer provides helps inform the audience; making the investors smarter about a particular security. <u>Well-presented</u> means that the article is well written; leveraging the right images, charts, data sources, general internet best use practices, laying the case for a stock out logically, easy to understand the thesis, good user experience*."

To gauge the validity of our three scores, $Score_{Convincing}$, $Score_{Actionable}$ and $Score_{Well\text{-}Presented}$, we select a random sample of 830 articles (5% of our total sample) and have them re-rated by 249 retail investors. In short, we find that the retail investor scores and SA editor scores are highly positively correlated, which helps build confidence in the validity of the SA editor scores.[11]

(g) Stock characteristics: Some types of stocks are likely more engaging to discuss than others. Holding all else equal, we suspect that an investment idea on an engaging stock is shared more frequently than an idea on a dull stock.

When gauging what types of stocks investors prefer conversing about, we do not deem it feasible to consider all possible firm characteristics. Instead, drawing from Han, Hirshleifer, and Walden's (2018), we test whether the characteristics that make a stock more engaging to discuss tend to be characteristics that put stocks in the short leg of anomalies whereas the characteristics that make a stock monotonous tend to be characteristics that put stocks in the long leg of anomalies.[12] That is, we test whether short-leg stocks are more engaging to converse about than long-leg stocks, which may help explain why the former (latter) tend to become overpriced (underpriced) in the first place.

To assess this conjecture, we construct *Short Score* and *Long Score*. These scores compute for how many out of 172 anomalies a stock resides in the short leg and the long leg, respectively. Short- and long-leg stocks are generally stocks in the top and bottom deciles (or quintiles) with respect to a given firm characteristic. Some firm characteristics are indicator variables and there is either only a long leg or only a short leg. For the full list of 172 anomalies and a description of each anomaly variable, please see Andrew Chen's website: https://sites.google.com/site/chenandrewy/Data. We thank Andrew Chen for graciously sharing his anomaly data with us.

[11] In particular, we design an online survey, in which we first present SA's explanations of their three scores, $Score_{Convincing}$, $Score_{Actionable}$ and $Score_{Well\text{-}Presented}$. We then show each survey participant ten SA articles. After reading each article, each participant is asked to assign scores indicating how "convincing," how "actionable," and how "well-presented" the article was found to be. Each article is rated by three participants. For each rated article, we take the average across the three individual scores. Online Appendix Figure 1 provides a screenshot of our online survey. We recruit survey participants via Prolific (https://prolific.ac). We require that participants' primary language be English and that they reply "Yes" to the following question: "Have you ever made investments (either personal or through your employment) in the common stock or shares of a company?"

[12] Han, Hirshleifer and Walden (2018) show that under the assumptions that senders report high past performance more often than low past performance and that receivers fail to adjust for this transmission bias and take past performance as indicative for future performance, stocks with high variance and high skewness are much more viral than stocks with low variance and low skewness. High variance and high skewness puts stocks in the short-leg of the volatility- and skewness anomaly.

As alluded to above, out of the seven features we consider in this study, three have been broached by prior marketing literature (sentiment, emotionality, usefulness). As we describe in the main body of the text, some of the insights based on consumer behavior extend to our investor setting; others do not and we provide possible explanations.

We include the following four control variables. Prior work suggests that word of mouth is stronger in the absence of complementary information sources (Berger 2014). The SA articles in our sample all reflect investment recommendations. The perhaps most significant alternate source of investment recommendations is professional sell-side analysts and the reports they compile. We thus measure the availability of complementary information sources through *ln (1 + Analyst Coverage)*, whereby *Analyst Coverage* is the number of analysts covering the stock in question. Our source for analyst data is the IBES database. Our second control is *Editor Pick*, which equals one if an SA article is chosen as one of the SA editors' most favored articles, and zero otherwise. All Editors' Pick articles are aggregated on a separate list on the SA website. They also receive an "Editors' Pick" tag near the title. We suspect that Editors' Pick articles are more frequently read to end and more frequently shared. Our third control is the natural logarithm of the number of words in a SA article, *ln (# words)*. We suspect that longer articles are less likely to be read to end. Finally, we include $ln(1 + \#\ Articles\ on\ Same\ Stock_{Previous\ Month})$, whereby $\#\ Articles\ on\ Same\ Stock_{Previous\ Month}$ represents the number of SA articles on a given stock over the previous month.

We provide descriptive statistics for the above variables in Table 1. Perhaps the most interesting descriptive statistic is that, on average, (only) 0.3% of the articles that are read to end also are shared via email. Such fraction may appear low. However, it is not unusual. For instance, in a case study, Twitter reports that only 0.07% of tweets seen by users are subsequently re-tweeted (Twitter 2014).

Importantly, the above fraction suggests that very few read-to-ends originate from individuals receiving an email about the article in question. Instead, almost all read-to-ends originate from "patients zero" who encounter the article in question through their own individual searches. That is, whatever sharing preferences we detect in this study are those of patients zero.

3. What Do Investors Like (Themselves) versus Like to Talk About with Others?

We begin our analysis with the following regression specification:

$$Y_{i,j} = \alpha + \textit{Article-Level Determinants}_i \beta + \textit{Stock-Level Determinants}_j \gamma + X_i \delta + \varepsilon_{i,j} \quad (1)$$

where i indexes an article on stock j. The dependent variable is the natural logarithm of either one plus the number of read-to-ends or one plus the number of times an article is shared via email. *Article-Level Determinants*$_i$ include *Sentiment*, *Extreme Sentiment*, *Unusual Sentiment*, *Emotionality*, *Numbers-versus-Text*, and $Score_{Convincing}$, $Score_{Actionable}$, $Score_{Well\text{-}Presented}$. *Stock-Level Determinants*$_j$ include *Short Score* and *Long Score*. X includes our controls. *T*-statistics are based on standard errors adjusted for heteroscedasticity and clustered by day of article publication. In additional tests, we include author-fixed effects and we find that our conclusions are largely unchanged (results available upon request).

We first discuss our results regarding features that grab investors' attention, tabulated in the first column of Table 2. All the determinants discussed below are statistically significant at either the 1% level or the 5% level.

Investment ideas are more frequently read to end if they are of more negative overall sentiment, greater uniqueness and greater emotionality. The estimate for *Sentiment* suggests that a one-standard-deviation decrease in sentiment leads to 2.7% more frequent read-to-ends; the two estimates for *Unusual Sentiment* suggest that a one-standard-deviation increase in uniqueness leads to 2.3% - 2.4% more frequent read-to-ends; the estimate for *Emotionality* suggests that a one-standard-deviation increase in emotionality is followed by 1.6% more frequent read-to-ends.

Articles with extreme views receive 7.8% fewer read-to-ends. Articles that contain many numbers are also less likely to be read to end. The estimate for *Numbers-versus-Text* suggests that a one-standard-deviation increase comes with 3.6% fewer read-to-ends.

Investors are more likely to finish reading an article if such article is more useful in both function and form. The estimates for $Score_{Convincing}$, $Score_{Actionable}$ and $Score_{Well\text{-}Presented}$ imply that a one-unit-increase in any of the above three scores is followed by 24.1%, 15.0%, and 5.3% more frequent read-to-ends. A growing literature in accounting and finance argues that, given the myriad of investment opportunities available, investors stop considering shares for investment and simply "move on" if the corresponding firm's disclosure documents are difficult to read (Lawrence 2013; Elliott, Rennekamp and White 2015; Hwang and Kim 2017). Our finding that poorly-presented articles are less frequently read to end provides direct evidence for this argument.

The estimates for *Short Score* and *Long Score* are both positive, suggesting that investors are more captivated by investment ideas about stocks that reside in either extreme of an anomalous firm characteristic. Our estimates indicate that residing in the short leg (long leg) of ten more anomalies is followed by 16.5% (5.7%) more frequent read-to-ends.

As expected, the estimates for our controls show that articles picked by editors receive a higher number of read-to-ends. Longer articles as well as articles written on stocks already covered by many professional sell-side analysts are less likely to be read to end.

Overall, we find that the investment ideas that captivate investors to a greater degree are of (1) more negative overall sentiment, (2) more moderate viewpoint, (3) greater uniqueness, (4) greater emotionality, (5) lower reliance on numbers, and (6) greater perceived usefulness. Investment ideas written on (7) stocks residing in either the long or the short leg of anomalies are also more captivating.

The above results shed light on how investors behave by themselves. Next, we consider how investors behave in a group and interact with one another. In column (2) we report the results obtained when regressing the number of shares on the same set of independent variables as in column (1) while controlling for the number of read-to-ends. These results indicate the degree to which specific features of an investment idea generate sharing conditional on an article having been read. In column (3) we report results obtained without controlling for the number of read-to-ends. These results may be interpreted as reflecting the product of (1) the level of attention an idea receives and (2) the frequency of sharing per one unit of attention that occurs.

While our previous results suggest that more negative investment ideas are more likely to be read to end, the results in columns (2) and (3) suggest that investors tend to keep such negative investment ideas mostly to themselves. It is the investment ideas of more positive overall sentiment that are shared more frequently and, consequently, become a greater part of investors' conversations. The estimate for *Sentiment* reported in column (3) suggests that a one-standard-deviation increase in sentiment is followed by 5.3% more frequent article shares. Our results are similar to Berger and Milkman (2012) who find that more positive *New York Times* articles are more likely to make the *New York Times'* most emailed list.

The estimates for *Emotionality* and *Numbers-versus-Text* also flip their signs, suggesting that while emotional and qualitative stock opinions are more frequently read to end, it is the impersonal and quantitative stock opinions that are shared more frequently. The estimates reported in column (3) suggest that a one-standard-deviation decrease in *Emotionality* is followed by 1.4% more article shares and that a one-standard-deviation increase in *Numbers-versus-Text* is followed by 2.8% more frequent article shares. Our emotionality result is different from Berger and Milkman (2012) who find that more emotional *New York Times* articles are more likely to make the *New York Times'* most emailed list. One possible explanation is that in financial markets emotional stock opinions are viewed as less balanced and convincing and thus are shared less frequently for impression-management-related reasons. In line with this view, we find that *Emotionality* and $Score_{Convincing}$ are strongly negatively correlated.

The estimate for *Short Score* remains positive. However, the sign of the estimate for *Long Score* flips from positive to negative. Our result is broadly consistent with the notion that stocks residing in the short leg of anomalies are highly engaging to discuss whereas stocks residing in the long leg do not make good conversation pieces. Our finding provides a partial explanation for why short-leg stocks tend to become overpriced while long-leg stocks tend to become underpriced.

To be clear, we do not claim that investors prefer conversing about short-leg stocks for each of the 172 anomalies we consider in this study. Our finding represents a mere average effect. For instance, one prominent outlier we detect in additional (untabulated) analyses is momentum: investors prefer conversing about winner stocks even though winner stocks reside in the long leg of the momentum anomaly.

The signs of the estimates for all other variables remain largely the same: The estimate for *Extreme Sentiment* is weakly positive in column (2); in column (3), it is negative and statistically significant. The estimates for $Unusual\ Sentiment_{Stock}$ and $Unusual\ Sentiment_{Author}$ remain strongly positive as do the estimates for $Score_{Convincing}$, $Score_{Actionable}$ and $Score_{Well\text{-}Presented}$.

Overall, our results regarding the determinants of an investment idea's virality suggest that investment ideas are shared more frequently if they are of the following types: (1) more positive, (2) more moderate viewpoint, (3) more unique, (4) more impersonal, (5) more quantitative and (6) more useful. (7) Investors enjoy conversing about stocks residing in the short leg of anomalies while stocks residing in the long leg of anomalies are considered monotonous. All in all, of the seven determinants of attention and virality we consider in this study, four have opposing effects (sentiment, emotionality, quantitative nature, short leg/long leg) while three have broadly similar effects (extremeness of viewpoint, uniqueness, usefulness). That is, what investors find captivating and what investors end up sharing with peers can be strikingly different.

4. What Are the Implications of Investors' Sharing preferences?

4.1 Implications for Investors' Trading Performances

It is possible that the abovementioned preferences and aversions represent what David Hirshleifer in his AFA 2020 presidential address refers to as a "social transmission bias": the types of investment ideas that investors prefer conversing about do not have higher investment value than the types of ideas that investors are averse to discuss. They are merely shared more frequently for impression-management- and emotion-regulation purposes. If receivers do not fully discount for this, sharing preferences may cause them to put too much weight on some pieces of information while ignoring other potentially useful insights.

The alternative perspective is that the types of ideas investors prefer sharing have greater investment value than the types of ideas investors choose not to share. That is, investors' preferences and aversions do not represent a social transmission bias but, instead, help filter out noise and efficiently aggregate value-relevant information.

To differentiate the bias view from the efficient-aggregation perspective, we build on Chen, De, Hu, and Hwang (2014), who provide evidence that the average SA article contains useful investment advice in the sense that the overall sentiment revealed in an SA article positively predicts the corresponding firm's subsequent stock market performance. We define the investment value of an idea as the degree to which the overall sentiment revealed in an SA article positively predicts subsequent returns and we estimate a regression of cumulative abnormal returns over one week, three months or six months following article publication on *Sentiment*, our determinants of virality, and *Sentiment* interacted with our determinants. The estimates for the interaction terms inform us whether the types of articles that investors prefer sharing have greater investment value than the types of articles that investors avoid sharing. Our regression also informs us which of the various sharing preferences are the potentially most beneficial or harmful to investors' trading performances. Following Chen, De, Hu, and Hwang (2014), when computing cumulative abnormal returns we skip the first two days of article publication and we compute abnormal returns as the difference between raw returns and returns on a value-weighted portfolio of firms of similar size, book-to-market ratios, and past returns (Daniel, Grinblatt, Titman, and Wermers 1997). *T*-statistics are based on standard errors adjusted for heteroscedasticity and clustered by stock and day of publication.

Since the results from the first part of our analysis suggest that the types of articles that investors, themselves, find the most captivating and the types of articles that investors end up sharing are often of diametrically opposed spectra in terms of content attributes, in Table 3, we also analyze whether it is (1) the types of articles with the more frequent read-to-ends or (2) the types of articles with the more frequent shares that have greater investment value. We estimate a regression equation of subsequent stock market performance on *Sentiment*, the natural logarithm of one plus the number of read-to-ends, the natural logarithm of one plus the number of shares, and *Sentiment* interacted with these two variables.

We discuss the results from this analysis first. As shown in columns (1), (3) and (5), we find that irrespective of the return horizon, the estimates for the interactions between *Sentiment* and *ln (1 + # Read-to-Ends)* are all positive and reliably different from zero, suggesting that the articles that are more frequently read to end have greater investment value. In stark contrast, the estimates for the interaction term with *ln (1 + # Shares)* are negative, albeit not statistically significantly so. That is, while the sentiment in articles that are more frequently read to end more accurately predicts future abnormal returns, the sentiment in articles that are shared more frequently does not. Put bluntly, individuals keep the more useful investment advice mostly to themselves.

As alluded to before, a primary motivation for individuals to engage in word-of-mouth communication is not so much to help their listeners, but to regulate their emotions and manage their impressions. We believe most of the features that we find encourage sharing can be seen as enhancing senders' desired impressions. It may thus not surprise that the ideas most frequently shared are not necessarily the ideas most useful to listeners' investment decisions.

In columns (2), (4), and (6), we explore which of our sharing preferences are responsible for the above performance wedge. The results are somewhat inconclusive as most of our estimates are statistically insignificant. However, we find hints in the data that the potentially most harmful sharing preferences are (1) investors' reluctance to share content that is mostly qualitative and more emotional and (2) investors' prepossession (disinterest) in discussing investment ideas on stocks residing in the short leg (long leg) of anomalies. In particular, while articles that are mostly qualitative, more emotional and pertinent to long-leg stocks are *less* likely to be shared, the results in Table 3 provide suggestive evidence that these are the articles of greatest investment value. In general, most of the interactions between sentiment and our sharing-preferences variables produce results, which suggest that investors' sharing behavior accentuates the *less* useful ideas and masks the more useful ideas.

In further tests, we experiment with an alternate empirical design. Most SA readers are likely retail investors. Retail investors rarely short (Barber and Odean 2007). SA readers should therefore be primarily interested in discussing stocks that are seemingly underpriced and represent attractive buying opportunities. Relatedly, Barber and Odean provide evidence that – regardless of whether through positive news or negative news – a stock entering retail investors' radar triggers buying activity. From these perspectives, the investment value of an SA article stems from its ability to predict *positive* abnormal returns.

We estimate a regression similar to regression equation (1):

$$Y_{i,j} = \alpha + \textit{Article-Level Determinants}_i \beta + \textit{Stock-Level Determinant}_j \gamma + X_i \delta + \varepsilon_{i,j} \qquad (2)$$

Y is now the cumulative abnormal returns over one week, three months or six months following article publication and we test whether the features that stimulate sharing (also) predict strong positive abnormal returns whereas the features that discourage sharing do not.

The results are reported in Online Appendix Table 1. Overall, when comparing the estimates reported in Table 2 with those reported in the Online Appendix, we find that there are more cases of sharing and abnormal returns going in the opposite (and "wrong") direction than cases in which sharing and abnormal returns go in the same (and "right") direction.

By and large, it seems to us that most of the sharing preferences that we consider in this study do not help investors make better investment decisions and are perhaps best described as a social transmission bias.

4.2 Asset Pricing Implications

Our final analysis considers whether the findings presented in the previous sections have asset-pricing implications. As alluded to earlier, a large body of work finds evidence that we can use various firm characteristics to help predict the cross-section of average stock returns. The standard procedure in this literature is to sort stocks each month based on a firm characteristic. The long and the short leg are typically defined as the extreme deciles or quintiles produced by such sorts; some firm characteristics are indicator variables and there is either only a long leg or only a short leg.

Studies in the literature find that the long leg outperforms the short leg and that the difference in performance is difficult to explain with traditional asset-pricing models such as the Capital Asset Pricing Model (CAPM) of Sharpe (1964) and Lintner (1965). One possible explanation for this finding is that the CAPM is the incorrect model (e.g. Hou, Xue, and Zhang 2015). Another possibility is that the seemingly anomalous returns are not robust (e.g. Harvey, Liu, and Zhu 2016). A third, "behavioral," explanation is that the anomalous patterns are robust and there is mispricing (e.g. Hirshleifer 2001; Barberis and Thaler 2003; Engelberg, McLean and Pontiff 2018; Chen and Zimmermann 2019). On this view, short-leg stocks carry characteristics that overly excite some investors. This causes some short-leg stocks to become overpriced. Long-leg stocks carry characteristics that do not appeal to investors. The corresponding lack of investor interest causes some such stocks to become underpriced.

Here, we adopt the behavioral view but speculate that the presence of a short-leg stock exhibiting exciting features is not a sufficient condition for overpricing. One needs both the presence of exciting features and investors sharing such features to observe the occasional overpricing. That is, if there are two stocks with the same firm characteristics that put them in the short leg for the same number of anomalies but the first stock is talked about frequently while the second is not, we predict that only the former may become overpriced even though, again, both stocks have the same *Short Score*.

Why would some short-leg stocks but not others be talked about? As of 2018, there are more than 4,000 domestic companies listed in the US and more than 21,000 US-registered investment companies to which investors can allocate money.[13] Even though our previous analysis shows that short-leg stocks are more likely to be talked about, presumably because they exhibit characteristics that make them more appealing for discussion, we still do not deem it plausible that investors can discuss all stocks residing in the short-leg of anomalies.[14]

To test our prediction, we again count for how many out of 172 anomalies a stock resides in the short leg, which we represent by *Short Score*, and for how many anomalies a stock resides in the long leg, which we represent by *Long Score*. We also compute, across all SA articles published on a given stock in a given month, the number of shares scaled by the number of read-to-ends. We construct *Low Virality*, which equals one if the above ratio is in the bottom 30% of its distribution, and zero otherwise.

Our first regression is a pooled regression of monthly stock returns on lagged realizations of *Short Score* and *Long Score* and year-month fixed effects. We cluster our standard errors at the year-month level. In line with findings reported in prior literature, the estimates in column (1) of Table 4 suggest that residing

[13] https://data.worldbank.org/indicator/CM.MKT.LDOM.NO?locations=US and https://www.ici.org/pdf/2019_factbook.pdf

[14] Our sharing data strongly corroborate the view that while investors prefer sharing ideas on short-leg stocks, they cannot plausibly share all stocks residing in the short-leg of anomalies (results available upon request).

in ten more short legs brings 1.1% lower returns a month, whereas residing in ten more long legs brings 1.2% higher returns a month.

Our main results start with column (2) of Table 4. In addition to *Short Score* and *Long Score*, we now include *Low Virality* as of the previous month along with interactions between *Low Virality* with *Short Score* and *Long Score*, respectively. The estimate for *Short Score* is -0.156 (*t*-statistic = -2.56) and the estimate for *Short Score* × *Low Virality* is 0.141 (*t*-statistic = 2.41). The estimates of opposing signs and similar magnitudes suggest that among stocks with *Low Virality* the predictability of *Short Score* drops to zero. That is, without sharing, a given stock experiences no abnormally low returns no matter in how many short legs such stock resides.

Since *Low Virality* is based on the ratio of the number of shares to the number of read-to-ends, the result presented in column (2) should not be confounded by the degree of attention a given short-leg stock receives. Nevertheless, in column (3) we directly control for attention by further including *Low Attention* as of the previous month and interactions of *Low Attention* with *Short Score* and *Long Score*, respectively. *Low Attention* equals one if the number of read-to-ends across all SA articles published on a given stock in a given month is in the bottom 30% of its distribution, and zero otherwise.

Column (3) shows that the estimate for *Short Score* × *Low Virality* remains largely unchanged (coefficient estimate = 0.138, *t*-statistic = 2.45). The estimate for *Short Score* × *Low Attention* is 0.102 (*t*-statistic = 1.25).

The insignificant effect of attention is in line with findings reported in the consumer behavior literature: The reason that word of mouth is presumed to have such powerful effects on consumer behavior (Berger 2014) and, by many practitioner accounts, is considered the most effective marketing strategy (Misner 1999) is that a recommendation from a trusted friend is up to 50 times more likely to trigger a purchase than a recommendation by an outsider and still more likely to trigger a purchase than a mere product mention or advertisement (Bughin, Doogan, and Vetvik 2010). Applying such reasoning to our study, one may speculate that what triggers a majority of investment decisions is not so much investors coming across an investment idea by themselves or hearing a recommendation from an outsider. What triggers a majority of investment decisions is hearing about an investment from a personal friend.

If virality creates overpricing, we should observe not only low subsequent returns as mispricing subsequently correct itself, but also high contemporaneous returns as the corresponding stock turns viral. Column (4) of Table 4 and Figure 1 test this assertion. In column (4), we estimate the same regression equation as in column (3), but consider the contemporaneous association between stock returns, *Short Score* and virality. In Figure 1, we estimate a regression of monthly stock returns (either in year-month t+1 or in year-month t) on the number of times a stock resides in the short leg of an anomaly and the number of times a stock resides in the long leg of an anomaly (as of year-month t). We do so separately for stocks accompanied with "High Virality" in year-month t and stocks accompanied with "Low Virality" in year-month t. High-virality stocks are stocks for which *Low Virality* equals zero; low-virality stocks are stocks for which *Low Virality* equals one. Figure 1 plots the coefficient estimates for *Short Score*, along with their corresponding 95% confidence intervals.

Panel A of Figure 1 shows that among stocks with low virality, irrespective of whether we regress returns in year-month t+1 or returns in year-month t on *Short Score*, the estimate for *Short Score* is essentially zero.

The results are strikingly different for stocks with high virality. Here, the estimate for *Short Score* is strongly positive for returns in year-month t, yet strongly negative for returns in year-month t+1. That is, residing in the short leg of more anomalies comes with substantially higher returns in year-month t and substantially lower returns in year-month t+1. These patterns are consistent with the idea that virality at time t creates overpricing over time t, which then subsequently reverses over time t+1.

In Panel B of Figure 1, we further separate high-virality stocks into those that are likely to be short-sale constrained and those that are unlikely to be short-sale constrained. We measure short-sale constraints via average daily lending fees (from Markit). High lending fees are indicative of high short-sale costs and high short-sale constraints. We consider stocks as likely (unlikely) to be short-sale constrained if their average daily lending fees are in the top (bottom) decile of their distribution.

In line with expectations, our results suggest that virality creates much greater overpricing over time *t*, which then reverses over time *t*+1, when stocks are short-sale constrained. Without short-sale constraints, there is little overpricing, even if virality is high.

We conclude this section with a couple of notes. Throughout all columns, the estimates for *Long Score* are strongly positive, while the estimate for the interaction of *Long Score* with *Low Virality* tend to not be reliably different from zero. The lack of significance for the interaction between *Long Score* and *Low Virality* is not unexpected. Many investors do not short and, instead, sit out of the market if they do not like a given stock (Chen, Hong and Stein 2002). This holds particularly true among retail investors (Barber and Odean 2007). Thus, whether many investors discuss a long-leg stock's many unappealing features or whether a long-leg stock is not discussed and does not enter investors' radar in the first place, the investor-behavior and, consequently, the asset-pricing implications are similar.

We would also like to stress that we do not believe that anomalous returns are generated by the sharing of SA articles per se. In this study, we merely assume that our results on the sharing behavior of SA readers provide a representative glimpse of the sharing behavior of non-negligible parts of the investor population. That is, we assume that when SA readers have a tendency to share ideas of certain types, so do non-negligible parts of the investor population.

5. Conclusion

In conclusion, we view our study as a building block for a new body of literature that focuses (1) on better understanding what select types of content investors enjoy (dislike) conversing about, (2) on learning whether such sharing preferences are better described as a social transmission bias or as a mechanism to aggregate information efficiently, and (3) on the overall implications of these factors for investor trading and asset prices.

Our study provides some initial answers to these questions. We document that investors have strong preferences for sharing certain types of content. We provide preliminary evidence that such sharing preferences may be best viewed as a social transmission bias and that sharing preferences can hurt investors' trading performances and lead to mispricing in the stock market.

Our first-stage results contain other possible implications that we do not test formally. For instance, our results show that positive views of a stock propagate substantially more than critical views, which makes it seem as though social finance may play a role in the formation of bubbles. Formally assessing this possibility along with developing and testing additional implications derivable from investors' sharing preferences should create a fruitful avenue for future research.

References

Amabile, Teresa M. "Brilliant but cruel: Perceptions of negative evaluators." *Journal of Experimental Social Psychology* 19, no. 2 (1983): 146-156.

Bali, Turan G., David A. Hirshleifer, Lin Peng and Yi Tang. “Attention, social interaction, and investor attraction to lottery stocks.” Georgetown University Working Paper (2019).

Banerjee, Abhijit, Arun G. Chandrasekhar, Esther Duflo, and Matthew O. Jackson. "The diffusion of microfinance." *Science* 341, no. 6144 (2013): 1236498.

Barber, Brad M., and Terrance Odean. "All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors." *The Review of Financial Studies* 21, no. 2 (2007): 785-818.

Barberis, Nicholas. "Psychology-Based Models of Asset Prices and Trading Volume." *Handbook of Behavioral Economics-Foundations and Applications* 1 (2018): 79-175.

Barberis, Nicholas, and Richard Thaler. "A survey of behavioral finance." *Handbook of the Economics of Finance* 1 (2003): 1053-1128.

Berger, Jonah. "Word of mouth and interpersonal communication: A review and directions for future research." *Journal of Consumer Psychology* 24, no. 4 (2014): 586-607.

Berger, Jonah. *Contagious: How to build word of mouth in the digital age*. Simon & Schuster, (2016).

Berger, Jonah, and Katherine L. Milkman. "What makes online content viral?" *Journal of Marketing Research* 49, no. 2 (2012): 192-205.

Brown, Jeffrey R., Zoran Ivković, Paul A. Smith, and Scott Weisbenner. "Neighbors matter: Causal community effects and stock market participation." *The Journal of Finance* 63, no. 3 (2008): 1509-1531.

Bughin, Jacques, Jonathan Doogan, and Ole Jorgen Vetvik. "A new way to measure word-of-mouth marketing." *McKinsey Quarterly* 2 (2010): 113-116.

Chen, Hailiang, Prabuddha De, Yu Jeffrey Hu, and Byoung-Hyoun Hwang. "Wisdom of crowds: The value of stock opinions transmitted through social media." *The Review of Financial Studies* 27, no. 5 (2014): 1367-1403.

Chen, Joseph, Harrison Hong, and Jeremy C. Stein. "Breadth of ownership and stock returns." *Journal of Financial Economics* 66, no. 2-3 (2002): 171-205.

Chen, Andrew Y. and Zimmermann, Tom. “Publication bias and the cross-section of stock returns.” Federal Reserve Board Working Paper (2019).

Daniel, Kent, Mark Grinblatt, Sheridan Titman, and Russ Wermers. "Measuring mutual fund performance with characteristic-based benchmarks." *The Journal of Finance* 52, no. 3 (1997): 1035-1058.

Dichter, Ernest. "How word-of-mouth advertising works." *Harvard Business Review* 44 (1966): 147-166.

Dubois, David, Andrea Bonezzi, and Matteo De Angelis. "Sharing with friends versus strangers: How interpersonal closeness influences word-of-mouth valence." *Journal of Marketing Research* 53, no. 5 (2016): 712-727.

Elliott, W. Brooke, Kristina M. Rennekamp, and Brian J. White. "Does concrete language in disclosures increase willingness to invest?" *Review of Accounting Studies* 20, no. 2 (2015): 839-865.

Engelberg, Joseph, R. David McLean, and Jeffrey Pontiff. "Anomalies and news." *The Journal of Finance* 73, no. 5 (2018): 1971-2001.

Forest, Amanda L., and Joanne V. Wood. "When social networking is not working: Individuals with low self-esteem recognize but do not reap the benefits of self-disclosure on Facebook." *Psychological Science* 23, no. 3 (2012): 295-302.

Fracassi, Cesare. "Corporate finance policies and social networks." *Management Science* 63, no. 8 (2017): 2420–2438.

Han, Bing, David Hirshleifer, and Johan Walden. “Social transmission bias and investor behavior.” University of Toronto Working Paper (2018).

Harvey, Campbell R., Yan Liu, and Heqing Zhu. "… and the cross-section of expected returns." *The Review of Financial Studies* 29, no. 1 (2016): 5-68.

Heimer, Rawley Z. "Peer pressure: Social interaction and the disposition effect." *The Review of Financial Studies* 29, no. 11 (2016): 3177-3209.

Heimer, Rawley, and David Simon. “Facebook finance: How social interaction propagates active investing.” Boston College Working Paper (2017).

Hirshleifer, David. "Investor psychology and asset pricing." *The Journal of Finance* 56, no. 4 (2001): 1533-1597.

Hirshleifer, David. "Behavioral finance." *Annual Review of Financial Economics* 7, (2015): 133-159.

Hirshleifer, David. "Social Transmission Bias in Economics and Finance." AFA 2020 Presidential Address.

Hong, Harrison, Jeffrey D. Kubik, and Jeremy C. Stein. "Social interaction and stock-market participation." *The Journal of Finance* 59, no. 1 (2004): 137-163.

Hou, Kewei, Chen Xue, and Lu Zhang. "Digesting anomalies: An investment approach." *The Review of Financial Studies* 28, no. 3 (2015): 650-705.

Huang, Shiyang and Hwang, Byoung-Hyoun and Lou, Dong. “What drives investor communication?” University of Hong Kong Working Paper (2019).

Hvide, Hans K., and Per Östberg. "Social interaction at work." *Journal of Financial Economics* 117, no. 3 (2015): 628-652.

Hwang, Byoung-Hyoun, and Hugh Hoikwang Kim. "It pays to write well." *Journal of Financial Economics* 124, no. 2 (2017): 373-394.

Kaustia, Markku, and Samuli Knüpfer. "Peer performance and stock market entry." *Journal of Financial Economics* 104, no. 2 (2012): 321-338.

Koetsenruijter, A. Willem M. “Using numbers in news increases story credibility.” *Newspaper Research Journal* 32, no. 2 (March 2011): 74–82.

Lawrence, Alastair. "Individual investors and financial disclosure." *Journal of Accounting and Economics* 56, no. 1 (2013): 130-147.

Lerner, Josh, and Ulrike Malmendier. "With a little help from my (random) friends: Success and failure in post-business school entrepreneurship." *The Review of Financial Studies* 26, no. 10 (2013): 2411-2452.

Lintner, John. "Security prices, risk, and maximal gains from diversification." *The Journal of Finance* 20, no. 4 (1965): 587-615.

Loughran, Tim, and Bill McDonald. "When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks." *The Journal of Finance* 66, no. 1 (2011): 35-65.

Lovett, Mitchell J., Renana Peres, and Ron Shachar. "On brands and word of mouth." *Journal of Marketing Research* 50, no. 4 (2013): 427-444.

Luminet IV, Olivier, Patrick Bouts, Frédérique Delie, Antony SR Manstead, and Bernard Rimé. "Social sharing of emotion following exposure to a negatively valenced situation." *Cognition & Emotion* 14, no. 5 (2000): 661-688.

Misner, Ivan R. *The World's Best Known Marketing Secret: Building Your Business with Word*, 2nd edition. Bard Press, (1999).

Peters, Kim, Yoshihisa Kashima, and Anna Clark. "Talking about others: Emotionality and the dissemination of social information." *European Journal of Social Psychology* 39, no. 2 (2009): 207-222.

Pool, Veronika K., Noah Stoffman, and Scott E. Yonker. "The people in your neighborhood: Social interactions and mutual fund portfolios." *The Journal of Finance* 70, no. 6 (2015): 2679-2732.

Sharpe, William F. "Capital asset prices: A theory of market equilibrium under conditions of risk." *The Journal of Finance* 19, no. 3 (1964): 425-442.

Shiller, Robert J., and John Pound. "Survey evidence on diffusion of interest and information among investors." *Journal of Economic Behavior & Organization* 12, no. 1 (1989): 47-66.

Shue, Kelly. "Executive networks and firm policies: Evidence from the random assignment of MBA peers." *Review of Financial Studies* 26, no. 6 (2013): 1401–1442.

Tesser, Abraham, and Sidney Rosen. "The reluctance to transmit bad news." *Advances in Experimental Social Psychology* 8, (1975): 193-232.

Twitter. "New Tweet activity dashboard offers richer analytics." (2014).

Vendemia, Megan A. "(Re) Viewing Reviews: Effects of Emotionality and Valence on Credibility Perceptions in Online Consumer Reviews." *Communication Research Reports* 34, no. 3 (2017): 230-238.

Table 1
Descriptive Statistics

In this table we present summary statistics for the variables in the main regressions. The study period runs from August 1, 2012 through March 31, 2013. *# Read-to-Ends* is the number of times SA users scroll to the bottom of an article. *# Shares* is the number of times an article is shared via email. *Sentiment* is the number of positive words minus the number of negative words divided by the total number of words in an article. *Extreme Sentiment* equals one if *Sentiment* is either in the top or bottom decile of its distribution, and zero otherwise. *Unusual Sentiment* $_{Stock}$ and *Unusual Sentiment* $_{Author}$ are constructed as follows: for each article written on stock *j* by author *k*, we construct the average *Sentiment* across all articles written on stock *j* over the previous month, *Sentiment* $_{Stock}$, and the average *Sentiment* across all articles composed by author *k* over the previous month, *Sentiment* $_{Author}$. Our measures of uniqueness, *Unusual Sentiment* $_{Stock}$ and *Unusual Sentiment* $_{Author}$, are the absolute differences between the sentiment of the article in question and its corresponding *Sentiment* $_{Stock}$ and *Sentiment* $_{Author}$, respectively. *Emotionality* is the number of positive words plus the number of negative words divided by the total number of words in an article. *Numbers-versus-Text* is the ratio of the total occurrences of numbers to the total number of words in an article. *Score* $_{Convincing}$ measures how convincing an article is, *Score* $_{Actionable}$ measures how actionable an article is, and *Score* $_{Well\text{-}Presented}$ measures how well-presented an article is. These three scores are all assigned by SA editors. *Analyst Coverage* is the number of analysts covering the stock in question. *Short Score* is the number of anomalies for which the relevant stock resides in the short leg. *Long Score* is the number of anomalies for which the relevant stock resides in the long leg. *Editor Pick* equals one if an article is selected as an "Editors' Pick," and zero otherwise. *# Words* is the number of words in an article. *# Articles on Same Stock* $_{Previous\ Month}$ is the number of articles about given stock over the previous month.

Variable	N	Mean	Std. Dev.	25^{th} Percentile	50^{th} Percentile	75^{th} Percentile
# Read-to-Ends	16,446	2,029.77	1,982.82	758	1,467	2,638
# Shares	16,446	5.46	8.51	1	3	7
Sentiment	16,446	0.00	0.01	-0.01	0.00	0.01
Extreme Sentiment	16,446	0.20	0.40	0.00	0.00	0.00
Unusual Sentiment $_{Stock}$	16,446	0.01	0.01	0.00	0.00	0.01
Unusual Sentiment $_{Author}$	16,446	0.01	0.01	0.00	0.00	0.01
Emotionality	16,446	0.03	0.01	0.02	0.03	0.03
Numbers-versus-Text	16,446	0.04	0.04	0.02	0.03	0.05
Score $_{Convincing}$	16,446	4.13	0.34	4	4	4
Score $_{Actionable}$	16,446	4.17	0.38	4	4	4
Score $_{Well\text{-}Presented}$	16,446	4.25	0.44	4	4	5
Short Score	16,446	14.96	8.34	9	13	19
Long Score	16,446	10.43	5.52	6	10	14

Table 1. Continued.

Variable	N	Mean	Std. Dev.	25th Percentile	50th Percentile	75th Percentile
Analyst Coverage	16,446	5.01	9.44	0	0	6
Editor Pick	16,446	0.08	0.28	0	0	0
# Words	16,446	1,000.38	621.13	620	866	1,187
# Articles on Same Stock $_{Previous\ Month}$	16,446	13.61	23.84	1	4	14

Table 2
Determinants of an Investment Idea's Attention and Virality

In this table we present coefficient estimates from regressions of the natural logarithm of one plus the number of read-to-ends and email shares that an article receives on characteristics of the corresponding article's content. The sample includes 16,446 opinion articles written on a single stock published on Seeking Alpha from August 2012 through March 2013. All variables are defined in Table 1. We do not report the intercept. *T*-statistics are reported in parentheses and are based on standard errors adjusted for heteroscedasticity and clustered by day of article publication. *, **, and *** denote statistical significance at the 10%, 5% and 1% levels, respectively.

	Attention	Virality	
	(1)	(2)	(3)
Sentiment	-2.385***	6.354***	4.784***
	(-4.18)	(14.62)	(8.41)
Extreme Sentiment	-0.078***	0.006	-0.046**
	(-4.28)	(0.33)	(-2.37)
Unusual Sentiment $_{Stock}$	3.235***	0.650	2.780***
	(3.03)	(0.77)	(2.57)
Unusual Sentiment $_{Author}$	3.270***	1.719*	3.872***
	(2.93)	(1.93)	(3.23)
Emotionality	1.542*	-2.337***	-1.323
	(1.88)	(-4.04)	(-1.63)
Numbers-versus-Text	-0.901***	1.304***	0.711***
	(-3.51)	(6.29)	(3.63)
Score $_{Convincing}$	0.241***	0.046*	0.204***
	(8.08)	(1.80)	(6.64)
Score $_{Actionable}$	0.150***	0.076***	0.175***
	(8.73)	(4.97)	(8.83)
Score $_{Well\text{-}Presented}$	0.053***	0.020	0.054***
	(3.48)	(1.41)	(3.37)
Short Score	0.017***	0.003***	0.014***
	(20.49)	(5.22)	(16.03)
Long Score	0.006***	-0.007***	-0.003***
	(4.85)	(-7.08)	(-2.48)
ln (1 + Analyst Coverage)	-0.091***	-0.028***	-0.088***
	(-11.90)	(-5.02)	(-13.17)
Editor Pick	0.114***	0.267***	0.342***
	(2.81)	(7.95)	(8.19)
ln (# Words)	-0.058***	0.381***	0.343***
	(-3.97)	(32.00)	(25.05)
ln (1 + # Articles on Same Stock $_{Previous\ Month}$*)*	0.266***	-0.106***	0.069***
	(45.03)	(-22.37)	(11.43)
ln (1 + # Read-to-Ends)		0.658***	
		(89.86)	
# Obs.	16,446	16,446	16,446
R^2	0.267	0.487	0.160

Table 3

Determinants of an Investment Idea's Attention and Virality: Social Transmission Bias or Efficient Aggregation of Information?

In this table we present coefficient estimates from regressions of cumulative DGTW-adjusted stock returns over one week, three months or six months following article publication (while skipping the first two trading days) on the overall sentiment expressed in an article, either all other independent variables from Table 2 or the two dependent variables from Table 2, and interactions between sentiment and those variables. We report only the estimates for the interaction terms with our main variables of interest. *T*-statistics are reported in parentheses and are based on standard errors adjusted for heteroscedasticity and clustered by firm and day of article publication. *, **, and *** denote statistical significance at the 10%, 5% and 1% levels, respectively.

	One Week		Three Months		Six Months	
	(1)	(2)	(3)	(4)	(5)	(6)
Sentiment × ln (1 + # Read-to-Ends)	0.135***		0.688*		1.142**	
	(2.68)		(1.93)		(2.51)	
Sentiment × ln (1 + # Shares)	-0.087		-0.140		-0.214	
	(-1.21)		(-0.58)		(-0.48)	
Sentiment × Extreme Sentiment		0.040		0.245		0.394
		(0.46)		(0.60)		(0.62)
Sentiment × Unusual Sentiment $_{Stock}$		-2.923		-0.289		-33.604
		(-0.64)		(-0.01)		(-1.06)
Sentiment × Unusual Sentiment $_{Author}$		-2.293		11.286		53.157**
		(-0.58)		(0.71)		(2.23)
Sentiment × Emotionality		1.449		9.284		12.634
		(0.26)		(0.53)		(0.48)
Sentiment × Numbers-versus-Text		-0.077		-4.997**		-9.541***
		(-0.14)		(-2.03)		(-2.64)
Sentiment × Score $_{Convincing}$		0.399		1.323**		0.891
		(1.23)		(1.99)		(0.96)
Sentiment × Score $_{Actionable}$		-0.258**		-0.758*		-0.803
		(-2.18)		(-1.77)		(-1.46)
Sentiment × Score $_{Well\text{-}Presented}$		0.339***		1.062**		1.529**
		(3.35)		(2.33)		(2.39)
Sentiment × Short Score		0.002		-0.069*		0.053
		(0.27)		(-1.69)		(0.73)
Sentiment × Long Score		0.003		0.087*		0.033
		(0.37)		(1.65)		(0.43)
# Obs.	13,427	13,427	13,427	13,427	13,427	13,427
R^2	0.001	0.001	0.011	0.017	0.012	0.028

Table 4
Investor Conversations and Anomalies

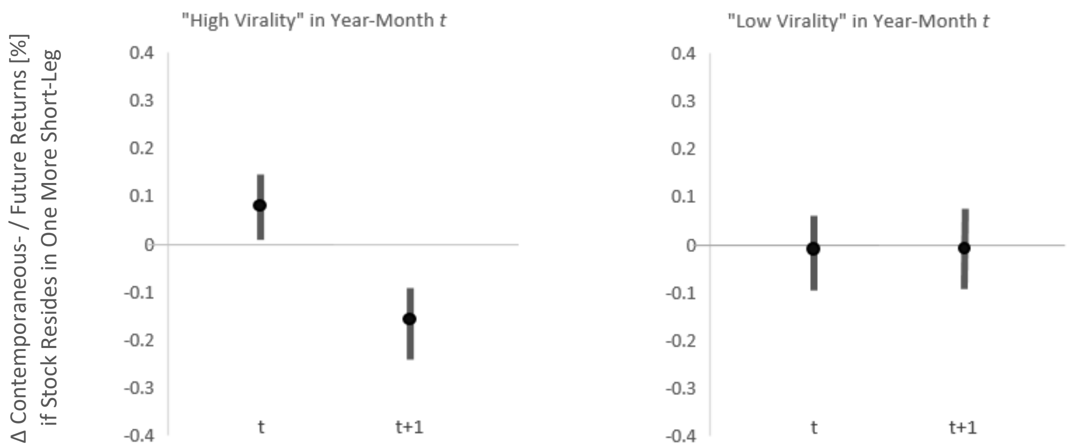
In this table we present coefficient estimates from regressions of monthly stock returns on the number of times a stock resides in the short leg or the long leg of an anomaly interacted with a measure of virality or attention. For each stock with an opinion article published on Seeking Alpha in year-month *t*, we compute its monthly stock return in year-month *t* and *t*+1. We also compute, across 172 anomalies, the number of anomalies for which a corresponding stock resides in the short leg (long leg), which we represent by *Short Score* (*Long Score*). Finally, for each stock, we compute the total number of times Seeking Alpha articles about the corresponding stock were shared through email (#*Shares*), scaled by the total number of read-to-ends (#*Read-to-Ends*). *Low Virality* is an indicator variable that equals one if #*Sharing* scaled by *#Read-to-Ends* is in the bottom 30% of its distribution and zero otherwise. *Low Attention* is an indicator variable that equals one if *#Read-to-Ends* is in the bottom 30% of its distribution an zero otherwise. We include year-month fixed effects. We multiply the coefficient estimates by one hundred. *T*-statistics are reported in parentheses and are based on standard errors adjusted for heteroscedasticity and clustered by year-month. *, **, and *** denote statistical significance at the 10%, 5% and 1% levels, respectively.

	Return (t + 1)			Return (t)
	(1)	(2)	(3)	(4)
Short Score	-0.110**	-0.156***	-0.172**	0.105**
	(-2.41)	(-2.56)	(-2.54)	(2.06)
Short Score × *Low Virality*		0.141**	0.138**	-0.104
		(2.41)	(2.45)	(-1.30)
Short Score × *Low Attention*			0.102	-0.027
			(1.25)	(-0.51)
Long Score	0.121**	0.170**	0.167***	0.118**
	(2.35)	(2.54)	(2.56)	(2.54)
Long Score × *Low Virality*		-0.165	-0.168	0.031
		(-1.57)	(-1.63)	(0.28)
Long Score × *Low Attention*			0.004	0.003
			(0.06)	(0.04)
Low Virality		0.187	0.238	0.514
		(0.14)	(0.18)	(0.82)
Low Attention			-0.927	1.350
			(-1.45)	(1.58)
# Obs.	4,314	4,314	4,314	4,314
R^2	0.058	0.061	0.062	0.056

Figure 1
Investor Conversations and Anomalies

This figure reports coefficient estimates from regressions of monthly stock returns (either in year-month *t*+1 or in year-month *t*) on the number of times a stock resides in the short leg of an anomaly and the number of times a stock resides in the long leg of an anomaly (as of year-month *t*). Here, we plot the coefficient estimates for *Short Score*, along with their corresponding 95% confidence intervals. In Panel A, we estimate our regression equation separately for stocks accompanied with "High Virality" in year-month *t* and stocks accompanied with "Low Virality" in year-month *t*. High-virality stocks are stocks for which the *Low Virality*-indicator variable (defined in Table 4) produces a realization of zero and low-virality stocks are stocks for which the *Low Virality*-indicator variable equals one. In Panel B, we further separate high-virality stocks into those for which the average daily lending fees in year-month *t* is in the top decile of its distribution ("high lending fees") and those for which the average daily lending fees is in the bottom decile of its distribution ("low lending fees").

Panel A. High- versus Low Virality



Panel B. High Virality and High- versus Low Lending Fees