eInformation: A Clinical Study of Investor Discussion and Sentiment

Sanjiv Das, Asís Martínez-Jerez, and Peter Tufano*

We examine the information flow for four stocks over seven months to trace the relationship between on-line discussion, news activity, and stock price movements. On-line discussions support numerous unsubstantiated rumors, substantial on-point exchanges, and quick dissemination of imminent and recently released information. Applying language-processing routines to message board postings and news, we create sentiment and disagreement measures or "eInformation." We analyze the determinants of sentiment and disagreement, and trace links between news, eInformation, and stock returns. This intensive clinical study of on-line discussions suggests mechanisms individual investors and groups can use to analyze and digest company information.

In light of the large body of research on informationally efficient markets, there seems little left to learn from the continued empirical examination of information and markets. It would seem similarly pointless for individual investors to try to compete with professional analysts. However, understanding the individuals' investing decisions has been one of the most vibrant research streams in recent years (see Roll, 1986, Odean, 1998, Gervais and Odean, 2001, and Barber and Odean, 2001).

Technology, in the form of stock chat message boards, now provides a new real-time window into discussions by individual investors. It is instructive to peek through this window to observe how information is digested, how sentiment evolves, and how perceptions are related to prices.

The method we adopt in this article is to use a clinical, i.e., small sample, approach to understanding investor behavior. Before framing hypotheses or constructing tests, it is important to establish a base level of understanding in an area. Thus, our article is decidedly descriptive, part of a long inductive tradition in economics (Blaug, 1992). We do not attempt to either affirm or reject theory. Rather, we suggest a series of working conjectures (or hypotheses) that can be developed through subsequent model building and large-scale empirical study.

We have three goals in this article. First, we closely analyze the people who share their opinions (posters) and their discussions surrounding a few stocks. Given the anonymous nature of this activity, we instead choose to study an outlier by interviewing an extensive poster. Doing so enables us to understand why someone would spend substantial amounts of time posting messages to one of the boards we study.

As part of our analysis, we also focus attention on the discussions themselves. Although

We would like to thank executives at The Motley Fool, Yahoo!, Silicon Investor, and Raging Bull for their cooperation; Sarah Woolverton, Sarah Erikson, John Kosmidis, and Jose Camacho for assistance with data collection; Mike Chen and John Sheridan for assistance with developing some of the computer code; and David Leinweber, Jacob Sisk, anonymous referees, and seminar participants at the HBS Brown Bag; the University of Texas at Austin, Yale, Ohio State, the Atlanta Finance Forum, the American Finance Association Meetings, and Northwestern University for their comments. We also thank Glenn, who candidly shared his experiences as a frequent poster. This research does not represent the views of any of the firms from which we obtained chat board information, nor the views of the firms studied. This work has been funded in part by the Dean Witter Foundation and a Breetwor Fellowship (Das), the Bank of Spain (Martínez-Jerez), and the Division of Research at the Harvard Business School (Tufano and Martínez-Jerez).

Sanjiv R. Das is an Associate Professor of Finance at Santa Clara University in Santa Clara, CA. F. Asís Martínez-Jerez is an Assistant Professor of Accounting and Control at Harvard Business School in Boston, MA. Peter Tufano is the Sylvan C. Coleman Professor of Financial Management in Boston, MA.

Financial Management • Autumn 2005 • pages 103 - 137

there is a perception that postings are "garbage," to the contrary, discussions sustain on-point exchanges, generate possibly non-public information, quickly disseminate public information from news stories, and serve as forums where investors can extract meaning from information. Chat rooms and postings are also sources of numerous unsubstantiated rumors, adding noise to the information flow. Nevertheless, the fact that even some non-public information may be released on the boards—and the observation that posters use the boards to test their own analyses and obtain those of others—may explain why posters and surfers continue to frequent these chat board sites.

Second, using language-processing algorithms, we measure the intensity and dispersion of sentiment (which we dub eInformation) for over 170,000 messages posted about four stocks. We analyze the determinants of the level of sentiment and disagreement among posters, and find that there is a close relationship between sentiment levels, stock prices, and trading volume. We also find that disagreement is related to the intensity of discussion.

Finally, we explore the usefulness of expressed investor sentiment (eInformation) to predict stock returns. Our clinical study confirms other studies that fail to find predictive power forecasting returns (Antweiler and Frank, 2002, 2004, and Das and Chen, 2003).

The article proceeds as follows. Section I deals with our clinical design. In Section II, we discuss the demographics of posters, detailing our interview with an especially active investor-discussant. Section III reports on our clinical examination of the nature of the discussions and the quality of information in those discussions. Section IV describes our computer-generated measures of sentiment and disagreement (eInformation) that are extracted using language-processing algorithms. In Section V, we analyze the determinants of our eInformation measures. In Section VI, we examine the relationship of eInformation to the price formation process. Finally, in Section VIII, we summarize the hypotheses that emerge from this clinical investigation.

I. Sample Design

We study four firms over a period of seven months. We use these four firms as archetypes for different information environments where traditional and new eInformation flows vary. As befits a clinical study, we attempt to dig deeply into these four firms, using our observations to derive hypotheses for large-scale studies. We have deliberately not selected pathological examples where posters have used stock message boards to explicitly manipulate prices (Leinweber and Madhavan, 2001).

To select the four firms for our preliminary study, we first collected information on the 3,724 firms that had at least one posting on The Motley Fool (TMF) stock message board during the period July 1, 1998 through January 31, 1999.[1] Then we classified each by the number of TMF messages. For the 504 firms in the TMF list that had at least 25 posts in the period July 1, 1998 through January 31, 1999, we collected the number of major news stories from Factiva. We defined a "major news story" as one in which the name of the company was either in the headline or was mentioned in the lead paragraph and appeared at least three times in the body of the article. We stratified the 504 firms into quintiles along two dimensions (number of posts and news stories) and selected one firm from each of the four extreme categories of the joint distribution. The four firms are shown in Table I. We did not select these four stocks to be representative of the average stock, but rather to help us to understand

[1]The Motley Fool graciously provided us with the data to perform this screening, but subsequent posting information for this and other boards was collected through a proprietary web crawler program.

Table I. Characteristics of the Four Companies Studied

This table provides the information based categorization, and basic business and financial information for the four companies in our study. The table gives both our data sources and dates. We obtain our information for this table from onesource.com, Hoovers, Bloomberg, and public filings. All dollar amounts are in millions. Financial figures are as of the end of each company's fiscal year: 12/31/98 for Amazon.com and General Magic, 6/30/99 for Delta Air Lines, and 3/31/1999 for Geoworks.

Panel A. Information Based Categorization of Companies

		Traditional Information Environment (number of news stories)	
		Rich (High)	Poor (Low)
eInformation Environment (number of posts)	Rich (High)	Amazon.com	General Magic
	Poor (Low)	Delta Air Lines	Geoworks

Panel B. Business and Financial Information

	Amazon.com	**Delta Air Lines**	**General Magic**	**Geoworks Corp.**
Business	On-line retailer	Major air carrier	Voice appl. service provider to telecom and Internet cos.	Provider of wireless software solutions
Industry	Retail (specialty; non-apparel)	Airlines	Software and programming	Communications services
Stock Listing (ticker)	NASD (AMZN)	NYSE (DAL)	NASD (GMGC)	NASD (GWRX)
Market Value (Year end)	$17054	$7984	$168	$56
Year Founded	1995	1924	1990	1983
Total Assets	$2471.6	$16750.0	$36.3	$18.2
Total Sales	$1639.8	$14597.0	$2.3	$8.8
Net Income	-$720	$1101	-$38.9	-$15.8
Institutional Ownership	30%	75%	10%	15%
Number of Institutions	442	776	61	54
Bond rating	B	BBB	not rated	not rated
Number of Analysts (equity + fixed income)	26 + 2	14 + 11	4+ 0	5+0
Number of Employees	2100	74000	169	110
Avg. Trading Volume (M shares / day)	27.75	1.19	1.04	0.28
Avg. Volume (% Outstanding)	9.37%	0.83%	3.50%	1.73%
Average $ Value of Trades/Day	$555	$65	$8	$1

the extremes of information flow (Table I, Panel A).

Table I, Panel B provides summary statistics on the four firms. Delta Airlines is an old-economy company with a large work force, substantial institutional ownership, and positive earnings. Amazon is considered a flagship new-economy company. General Magic and Geoworks are small, not very profitable, firms attempting to serve the new economy, but General Magic—founded by former Apple Computer executives—has an extremely high level of posting activity for a firm its size. Each of these firms is unique, and each is characterized by quite different information flows, which is the dimension along which we stratify this sample.

Table I shows that these firms are different along other dimensions as well, particularly in levels of trading activity. The two firms with substantial discussion are those with active trading (Amazon and General Magic with 9.4% and 3.5% daily turnover, respectively) and the firms with less substantial discussion show less active daily trading (Delta's trading volume in shares and value exceeds General Magic, but its share turnover is a quarter of General Magic's).

Our sample period is characterized by a high degree of uncertainty about the future, during which perception and sentiment drove values. It is therefore a particularly appropriate period for this study.

II. The Nature of the Posters

Posting a message is a quasi-anonymous act. Posters select a screen-name (or several screen names) and select how much information to reveal. Posters may not be "representative" of the average or marginal investor. For Delta, where institutional investors hold 75% of the shares, posters—who by all accounts are individual investors—surely do not represent the average investor. For the smaller stocks (General Magic and Geoworks), where individuals hold 85% to 90% of all shares, posters are likely to be more representative.

A. Number of Posters and Posting Activity per Screen Name

Table II reports data on the posters for our four boards for the stocks we study. For example, over our study period, 12,169 unique screen-names post 102,820 messages on the four Amazon boards. Although some individuals might have posted under multiple names, we suspect that most of these names represent unique individuals. For the other three stocks, the number of unique posters ranges from 404 (for Delta) to 3,208 (for General Magic). We can compare these numbers with the number of holders of record for each of these stocks in our sample period. If all of the posters are investors, they would represent 2% of the registered holders of Delta, but 528% of the shareholders of Amazon.[2]

Using traditional measures, the boards do not appear concentrated. In Table II, Panel A, we calculate Herfindahl indices for the 16 stock boards (four stocks times four board vendors). In only four cases is the share of message concentration at about the level that the Department of Justice would consider "mildly concentrated" in product markets (i.e., 1,000), and in each the number of postings is small. However, the distribution of posting activity is highly skewed. Table II, Panel B reports the number of postings by screen name. There is a relatively small and vigorous core of frequent posters, surrounded by a large number of occasional posters and by unobserved "lurkers," who only read the postings.

While we might understand why someone might post a few messages and then lose interest, it is less clear what motivates someone to post over 5000 messages about a single stock in a bit more than half a year. The time and effort expended by this person was considerable. Why?

[2] As of fiscal year end 1998 the number of registered shareholders for the four sample firms were: AMZN (2,304); DAL (21,672); GMGC (725); GWRX (7,800). Source: Compustat. The small number of AMZN shareholders likely reflects individual investors holding shares through omnibus brokerage accounts.

Table II. Posting Activity by Screen-Name and Poster Concentration

This table provides information on posting activity per poster on the four major stock message boards (Yahoo!, The Motley Fool, Silicon Investor, and Raging Bull) for the period July 1, 1998 through January 31, 1999 for the four stocks. A poster is defined by a unique screen name. Panel A uses the Herfindahl measure:

$$\sum_{i=1}^{n}\left(Marketshare_i * 100\right)^2$$

where the market-share of poster i is the share of messages over the period we study:

$$Numberofpostings_i \,/ \sum_{j=1}^{n} Numberofpostings_j \,.$$

Panel A. Message Share by Boards: Herfindahl Indices

Stock	Board	Number of Posters	Herfindahl Index	Number of Postings
AMZN	Raging Bull	162	286	517
	Silicon Investor	866	478	29543
	TMF	1031	175	10854
	Yahoo	10110	25	61906
	Average	3042	241	25705
DAL	Raging Bull	0	n/a	0
	Silicon Investor	10	2648	47
	TMF	23	510	31
	Yahoo	371	171	1313
	Average	101	1110	348
GMGC	Raging Bull	43	842	238
	Silicon Investor	189	214	2297
	TMF	62	455	185
	Yahoo	2914	60	62164
	Average	802	393	16221
GWRX	Raging Bull	6	1837	7
	Silicon Investor	29	1514	172
	TMF	10	2152	29
	Yahoo	359	274	1764
	Average	101	1444	493

B. Profile of an Active Poster

In a large empirical study, it is normal to discard anomalous observations. In contrast, in a clinical study, we can and should understand outliers. It is in this spirit that we interviewed Glenn R., the most prolific poster on the Amazon boards. Glenn gave an interesting interpretation of membership in a posting group, which we report at length to give readers a first-hand look at an active poster.

Table II. Posting Activity by Screen-Name and Poster Concentration (*Continued*)

Panel B. Number of Messages Posted over the Sample Period by Screen Name and Stock

	AMZN		DAL		GMGC		GWRX		TOTAL	
No. of Postings	**No. of Posters**	**%of Posters**	**No. of Posters**	**% of Posters**	**No. of Posters**	**% of Posters**	**No. of Posters**	**% of Posters**	**No. of Posters**	**% of Posters**
>4,000	1	0%	-	-	-	-	-	-	1	0%
1,001-4,000	7	0%	-	-	6	0%	-	-	13	0%
501-1000	10	0%	-	-	14	0%	-	-	24	0%
101-500	114	1%	-	-	101	3%	2	0%	217	1%
51-100	135	1%	3	1%	112	3%	5	1%	255	2%
26-50	281	2%	5	1%	150	5%	6	1%	442	3%
11-25	730	6%	17	4%	349	11%	14	3%	1,110	7%
6-10	949	8%	27	7%	365	11%	37	9%	1,378	9%
2-5	3,964	33%	111	27%	992	31%	126	31%	5,193	32%
1	5,978	49%	241	60%	1,119	35%	214	53%	7,552	47%
Total	12,169	100%	404	100%	3,208	100%	404	100%	16,185	100%

Glenn was in his late 40s when he was posting the messages we studied. He has an undergraduate degree in engineering from a large Midwestern university, and had completed some of the requirements for a business degree. He owned a small chain of jewelry stores, including an on-line jewelry store, and was self-employed. He did most of his postings on nights and weekends, when he was not otherwise busy at work. He estimated that he spent approximately 30 hours a week interacting on the boards.

He was a client of a large brokerage firm and read the professional analyst reports he received. A few years later, when we interviewed him, he was still able to cite analysts by name. He was interested in stocks and in technology, so he gravitated toward tech stocks. He also actively searched the web for news stories about these stocks. He provided four explanations for his activity.

1. Learning

"I wanted to learn." Glenn repeatedly emphasized that he lived in a small town of 15,000 people and that there were no investment clubs in his town. He reports, there were "not many people in town that he could talk to about investing." His activity in the Silicon Investor board was equivalent to membership in an investing club. Glenn was keen on learning from "people who had more experience than (he) had." In particular, he felt the boards were quite good in providing information on market microstructure details and technical analysis, especially the nuances of shorting stocks and the daily fluctuations in the outstanding float of the stock. Glenn approached stocks from the perspective of fundamental analysis, but was intrigued by the approaches of technical analysts that seemed to give them "a better batting average."

2. Complementing Professional Analysts

He felt that the professional analysts missed many of the details about firms, and he used the discussion boards to test out his analyses. Glenn did not believe that he, or any of the active members of the Amazon board, had any proprietary or inside information. "I don't think there was any truly inside information…the whole group had no better idea than the next person." However, they did have the time, experience, and inclination to carefully analyze the fundamental data on Amazon. As he explains, "I was perceiving this firm as a retailer and I was in the retail business. There was no question that the cost of fulfillment was higher than in regular stores. Others didn't understand issues of costs." Although much of this information was in public disclosures, it was buried in footnotes and labor intensive to pull out. This information was "missed by a lot of the analysts."

3. Interaction with Colleagues

The boards provided Glenn with colleagues that he enjoyed. We observed 925 posters on Silicon Investor during our study period, but Glenn estimated that there was a much smaller number (50 or 60) that were relatively active. Of these he came to know five or six personally, through phone calls or in-person meetings. Unlike the "cheerleaders," these people helped each other "see through" the news stories. They discussed stock picks and non-investing business advice off-line.

4. Self-Esteem

The boards provided Glenn with a venue to engage in enjoyable debate and to earn the respect of others. Glenn called this interaction the "entertainment value" of the boards, the ability to engage others in sustained discussion. Moreover, the discussion was self-reinforcing. “I enjoyed putting forth an opinion and then having to justify it.” According to Glenn, people who earned positive reputations were those who were able to more accurately predict the short-run stock price or the next earnings numbers, and those who provided superior insights. Glenn was proud to develop a reputation for the latter. “People wanted to know what I thought…it was a feeling of accomplishment.”

In retrospect, Glenn felt that he lost more money as a result of participating on the boards than if he had not. By virtue of having to stake out and argue a position in public, he felt that he probably became more “stubborn” about his opinions, and held onto his positions longer than he might have otherwise. He reports that while he lost money on his Amazon position, he profited on a few other positions that he followed regularly.

This interview gives a new dimension on discussion boards and the investing process. It reminds us that investing is not necessarily a solitary activity, but can be a communal activity (Das and Sisk, 2005). People voluntarily join communities because they perceive some benefits. For Glenn, the benefits included the enjoyment of coming to know other like-minded people, the ability to share ideas, and the ability to develop a reputation for clear thinking. An on-line community focused on a particular stock is no less valuable to its participants than one that revolves around a television show or game.

These observations provide motivations for voluntary postings on stock chat message boards. The theoretical justification for posting might be related to models of information disclosure. Suppose each investor receives a noisy signal about future stock price, e.g., their opinion as to the importance of a new product announcement. By sharing their signals with others, they can verify the information before trading, or can share the signal with others after trading, with the hope that their interpretation will lead to the desired movement in share prices. On a more mundane level, stock chat boards can be locations where disgruntled shareholders, customers, employees, and former employees can share their experiences with others.

To the extent that the online community serves as a social group or debating society, its economic impact is probably secondary. However, to the extent that it serves as a vehicle for testing ideas and analyses, it frames some interesting questions that could form the basis for subsequent research. We could ask, what are the relative returns from communal compared to individual analysis? Glenn believed that his analysis would be improved by testing it with others, but this is an untested assertion. More narrowly, can discussion, even among well-meaning investors, have the impact of producing even more severe biases, like the hardening of Glenn’s investment bias? Ex post, Glenn reached this self-critical conclusion, but it could be a more general phenomenon.

III. The Substance of the Discussions

Exploiting our clinical research design, we analyze the content of the postings. We look at what subjects are discussed, whether discussions stay on point, and whether the discussions reveal meaningful information.

We conduct two analyses. In the first, we examine the actual news released by firms, and look before and after the news release to understand any foreshadowing of the news prior to the release and the subsequent digestion of the news after the release. From this analysis we find that the discussion boards seem to play an important role in rapidly disseminating news, sometimes “breaking stories” before they are covered widely. In the second analysis, we examine a set of rumors on the boards and track them through time. From this analysis, we conclude that discussion boards are rumor mills for many unsubstantiated claims and a poor source of inside information.

In the first experiment, we selected 16 seemingly-newsworthy press releases by the four companies, based on our reading of the releases and inspection of abnormal returns around the announcements. We performed event studies on the 16 announcements and found that ten of the events have abnormal 1-day or 3-day returns over 5%, however given the high level of volatility in general, these returns are statistically significant in only four cases. For each of these companies we trace how the "news" is communicated to investors through traditional media as well as through postings (Table III lists the events). We also try to understand how the press or message boards provided advance information of the event, and how the companies responded to the event.

To analyze response, we measured the speed from the press release to the first discussion on the message boards, the time series of subsequent discussion, and the nature of the discussion. We find the following patterns, which we think of as empirically driven "hypotheses" about the different functions of the boards:

A. Message Boards Provided Factual Foreshadowing of Subsequent Press Releases

In quite a few instances, posters provided readers with advance warning of subsequent news events. For example, one of our tracer events is General Magic's spin-off of its DataRover division. Nine days before the DataRover spin-off, someone reported on the Yahoo! message board that they had found a new DataRover website that did not mention General Magic. This site was apparently taken off-line in a few hours by General Magic, which was probably testing the URL for the imminent spin-off. The board's readers could not only read the message, but also confirm it by going to the site.

In another example, one day before General Magic announced an agreement with Microsoft, someone posted that the two looked like they would share a booth at the Consumer Electronics Show, a tip-off to some closer relationship. A third General Magic poster alerted readers to a local radio broadcast that had suggested that the firm would enter into an agreement with Intuit. The agreement was not publicly announced until a few hours later.

In other instances, posters speculated about upcoming stock splits and bond issues. In two others, we see advance discussion of upcoming earnings numbers, or so-called "whisper numbers" as studied by Bagnoli, Beneish, and Watts (1999).

This anecdotal evidence suggests that posters provide active surveillance, especially of smaller companies, well before the traditional press picks up news events. In these instances, the posters seemed to "stumble across" non-public information, rather than being privy to inside information. Having stumbled across it, they then shared it, possibly to get others to validate it or to make sense of it.

B. Rapid Postings Disseminate Company Information Quickly

Companies tend to issue press releases either before markets open (sometimes in the middle of the night) or after markets close. Table III shows the number of minutes between the time-stamps on each press release and the first posting of the news on one of the stock chat boards. For many, but not all, announcements, the first post is incredibly soon after the news is posted, and in a few cases, prior to the time stamp of the first major news wire story.

Table III. Tracer Analysis: Backwards Analysis of Subsequently Released News Stories

We report on 16 press releases made by the four companies that seem to indicate a newsworthy change in the firm's business or financing. For each event, we summarize the speed of information dissemination by traditional and chat room sources.

Company	Event	Release Time	First Press Notice	Mins. to First Press	Related Press Activity	First Posting	Mins. to First Post	Subsequent Related Postings	Foreshadowing
Geoworks	Poor 1st quarter results	1/26/99 8:02 a.m. EST	Dow Jones NewsWire 1/26/99 9:34 a.m. EST	92	Warning of Price drop 1/25/99	1/26/99 8:45 a.m (Yahoo!)	43	15 before, 8 after market close.	Disc. of poor earnings possibility given 1/25/99 price drop
Geoworks	GWRX announces cooperation with Optimay	1/20/99 6:00 a.m. EST	Dow Jones NewsWire 1/21 3:53 a.m. EST	1,313	Warning of Price Jump 1/19/99	1/20/99 7:07 a.m. (Yahoo!)	67	18 before, 9 after market close; 4 next trading day.	—
Geoworks	Dave Grannan named as CEO	1/11/99 8:05 a.m. EST	Dow Jones NewsWire 1/11 8:03 a.m. EST (prelim.)	-2	1/11: 2 Dow Jones Articles 1/12: Wall Street Journal	1/11/99 9:15 a.m. (Yahoo!)	70	16 before, 4 after market close; 6 next trading day.	—
Geoworks	Debuts enhanced phone with Mitsubishi	11/16/98 6:00 a.m. EST	Dow Jones NewsWire 11/16 11:10 a.m. EST (preliminary)	310	11/16: 7 Dow Jones Articles, 4 Price Alerts; 1/17 Wall Street Journal	11/16/99 10:06 a.m. (Yahoo!)	246	35 before, 80 after market close.	—
Delta	Alliance with Korean Air	8/6/98 8:00 p.m. EST	Dow Jones NewsWire 8/6 8:25 p.m. EST	25	8/6: 1 Dow Jones Article; 8/10: M2 PressWire release	—	never	—	—
Delta	Finances beat consensus	7/16/98 9:14 a.m. EST	Dow Jones NewsWire 7/16/98 8:59 a.m. EST (prelim.)	-15	7/16: 4 Dow Jones Articles; 7/17: 4 newspaper, 1 BusinessWire	7/16/99 9:25 a.m. (Yahoo!)	11	2 before market close	—

Table III. Tracer Analysis: Backwards Analysis of Subsequently Released News Stories (*Continued*)

Company	Event	Release Time	First Press Notice	Mins. to First Press	Related Press Activity	First Posting	Mins. to First Post	Subsequent Related Postings	Foreshadowing
Delta	Announces strategic mgmt. reorg. in int'l markets	12/9/98 7:01 a.m. EST	Dow Jones NewsWire 12/9 7:03 a.m. EST (prelim.)	2	12/9: 3 Dow Jones Articles; 12/10: M2 PressWire	—	never	—	—
Delta	2 for 1 stock split approved	10/22/98 1:54 p.m. EST	Dow Jones NewsWire 10/22 1:55 p.m. EST (prelim.)	1	10/22: 2 Dow Jones articles; 10/23: M2 PressWire	10/22/98 4:01 p.m. (Yahoo!)	127	1 after market close; 3 on 10/23/98	Questions about split timing 10:46 a.m.
General Magic	Agreement with Intuit for voice access to financial info	11/9/98 4:04 a.m. EST	Dow Jones NewsWire 11/9 4:09 a.m. EST (prelim.)	5	11/9: 4 Dow Jones Articles; 11/10: 1 Dow Jones Article	11/9/98 12:39 a.m. (Yahoo!)	515	54 before market open; 112 before, 90 after market close.	Info of radio broadcast of Quicken news before press release.
General Magic	2nd quarter results	7/29/98 4:02 p.m. EST	Dow Jones NewsWire 7/29 4:06 p.m. EST (prelim.)	4	7/29: 2 (Dow Jones, BusinessWire) 7/30: Wall Street Journal	7/29/98 4:11 p.m. (Yahoo!)	9	130 between release and 9 p.m. that night.	Discussion of potential results beforehand
General Magic	Alliance with Microsoft for auto-enabled PC	1/7/99 7:34 a.m. EST	Dow Jones NewsWire 1/7 2:23 p.m. EST (prelim.)	409	1/7: 2 Dow Jones Articles	1/7/99 7:53 a.m. (Yahoo!)	19	59 before market open; 179 before market close.	Rumor of shared booth at CES between MSFT and GMGC (1/6 10:12 a.m.)
General Magic	Spin off of DataRover division as independent company	10/28/98 8:10 p.m. EST	LA Times and New York Times, 10/30	>1500	10/30: 3 (Bloomberg, ComputerWire, LA Times) 11/2: 1 Electronics News (2 articles on Q3 results)	10/28/98 8:41 p.m. (Yahoo!)	31	32 between release and market open on next trading day.	Posting of suspicious independent DataRover URL on 10/19/98.

Table III. Tracer Analysis: Backwards Analysis of Subsequently Released News Stories (*Continued*)

Company	Event	Release Time	First Press Notice	Mins. to First Press	Related Press Activity	First Posting	Mins. to First Post	Subsequent Related Postings	Foreshadowing
Amazon	Acquisition of Junglee	8/4/98 7:30 a.m. EST Note: 8-K filed 8/8/98	Dow Jones NewsWire 8/4; 9:05 a.m. EST	95	8/8: 17 8/9: 69 8/10: 3 8/11-18: 27	8/4/98 7:47 a.m. (Silicon Investor). 7:58 a.m. (Yahoo!)	17	5 before market open; 21 before, 22 after market close.	—
Amazon	$500 million debt issue	1/28/99 7:25 a.m. EST	Dow Jones Newswire 1/28 7:41 a.m. EST	16	1/29: 1 Dow Jones Article 1/29: 6 2/1: 4	1/28/99 7:40 a.m. (Yahoo!)	15	34 before market open; over 300 more before market close.	Bond discussion in days before (no facts, though)
Amazon	Enters European book market	10/15/98 5:00 a.m. EST	Dow Jones Newswire 10/15 5:51 a.m. EST	51	10/17: 3 Newspaper Articles	10/15/99 6:10 a.m. (Motley Fool)	70	6 before market open, 6 before close.	—
Amazon	Amazon announces 3:1 Split	11/19/98 5:45 p.m. EST	Dow Jones NewsWire 11/19 5:47 p.m. EST	2	11/20: 9 11/21: 3 11/22: 1 11/23: 1	11/19/98 5:51 p.m. (Yahoo!)	6	Over 300 before midnight that day.	Split speculation in days before announcement

First posts tend to contain a short notice of the news, often with a URL directing readers to the press release. The boards are apparently serving to disseminate information to interested investors quite quickly. This observation is consistent with those of the poster, Glenn, who mentioned that it was his routine to scan the press about Amazon and post links to new and important stories.

C. Extensive On-Point Discussion Is Sustained for Eight Hours After News Releases

Figure 1 shows the postings that followed these 16 events. Panel A shows that in the few hours immediately after a news event, posting volume rises, but then tails off over time. Panel B displays the composition of the posts over the first eight post-news hours.

We measure the nature of the discussion by categorizing each subsequent post into one of five possible categories (asks question, offers alleged fact, shares opinion, comments unrelated to news event, and spam/garbage). The first three categories are on-point postings, i.e., ones that relate to the news at hand. We see that for the first four hours after a news event, over two-thirds of all posts are on-point, and even eight hours later, about half are still discussing the news (as opposed to other issues or spam).

D. The On-Line Discussion Is a Mix of Questions, Answers, and Opinions

We categorize on-point posts as asking a question, offering an alleged fact, or proposing an opinion about the meaning of the news. For the first hour, we see more of a question-and-answer pattern, with a quarter of all posts and a third of the on-point posts either asking a question or supplying a fact. Over time, the discussion tends toward more analysis, i.e., interpreting facts, an observation that is consistent with our understanding of the primary function of the board and with our discussions with Glenn, the prolific Amazon poster.

The research design in this first experiment is biased in that it is conditioned on validated news, i.e., all of the stories we studied were real.

In the second analysis, we note and track rumors on the boards. In particular, we search our sample postings for messages related to mergers and acquisitions, which are material corporate events. The keywords we use for the search are merge, merger, hostile, acquisition, acquire, takeover, target, tender, offer, and stock swap. Our goal is to identify events with enough potential materiality that they might be ultimately reported.

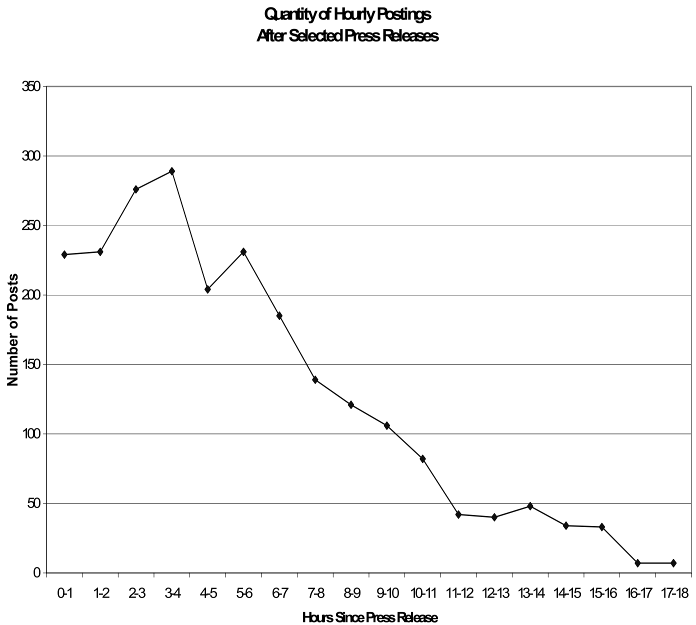
We also searched for news related to the posting rumors in Factiva for the period January 1998 to August 1999 (i.e. six months before and after our sample period). In total, we identified 54 merger and acquisition rumors on the discussion boards, with seven of these meriting at least five posts. However, the rumors are almost entirely ungrounded in fact. In two instances, the press story announcing the rumor cited the Internet as the source of the rumors.

Just one of the 54 rumors preceded an announcement by the company of an actual merger. Nine others preceded similar rumors in the business press, but not an actual transaction. One preceded a press denial by the company. The majority (43 of 54) did not result in a transaction or even in a rumor in the press. The remaining observation is a rumor on the boards that followed, rather than preceded, some press rumors.

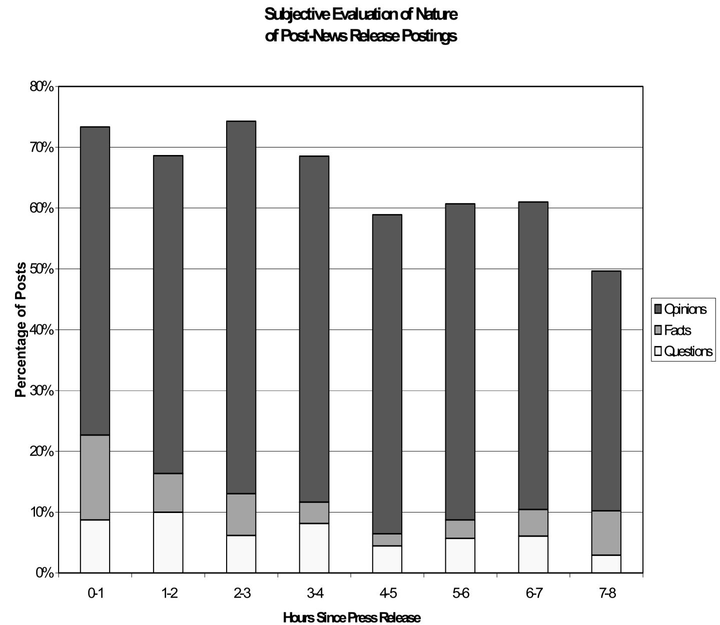
This second analysis produces a different picture from the first. Boards quickly disseminate information and provide investors with a forum to digest it. They also share not-quite-yet-public information. But in our sample, the boards do not seem privy to truly material inside information. This observation is consistent with Glenn's observations and the empirical

Figure 1. Discussion Activity Around 16 News Events

We report on 16 press releases made by the four companies that seem to indicate a material change in the firm's business or financing. For each event, we summarize the speed of information dissemination by traditional and chat room sources. Panel A shows the aggregate number of posts after the time-stamp of the press release. For each news release, we manually categorize the postings in the subsequent eight hours into on-point (relating to the news event) or off-point (unrelated to the news event or spam). We also categorize the on-point postings into ones that ask a question, offer an alleged fact, and propose an opinion. Panel B shows the distribution of the on-point postings. Table III provides detailed information on the nature of the events, the first posting activity, and the first press activity.



Panel A. Number of postings by hour after 16 selected corporate press releases.



Panel B. Distribution of type of posting by hour after 16 selected corporate press releases. Postings are classified as on-point if related to the news story, and off-point otherwise. The histogram shows the percentage of on-point posts (the height of each bar) and the nature of the on-point posts (asks question, provides alleged fact, proposes opinion.)

results on return predictability, which we discuss later.

There is a less obvious conclusion from this analysis. Although the boards are places where many rumors are suggested, we did not see evidence that they were "rumor mills," where these rumors themselves were the source of sustained discussion. In 47 of 54 instances, each of the rumors generated fewer than four subsequent posts, and unsubstantiated rumors generated less discussion.

If we were to construct a "wheat and chaff" measure for the boards, they would probably perform poorly. On the positive side, of the 16 actual news events in Table III, half were foreshadowed on the board discussions. This average sounds good, but this is conditional on knowing that something actually happened. In contrast, an avid reader of the boards continually scanning for merger announcements would have had useful information 2% of the time, with the remainder of the stories being unsubstantiated rumors. Although it is impossible to compare these percentages without knowing the gains and losses of trading on this information, it seems that the gains from being right 2% of the time would be more than offset by being wrong 98% of the time.

IV. The Concept and Measurement of eInformation

In the remainder of the study, we use computer algorithms to classify the 170,953 messages and to relate our measures of sentiment and disagreement to both information sources and stock prices.

A. Definition and Motivation of eInformation

The simplest characterizations of the flow of information are activity measures: simple counts of the numbers of news stories or postings, or the length of news story or posting. These metrics are used by Mitchell and Mulherin (1994; number of news stories) and Wysocki (1999; number of postings). These activity measures indicate the level of interest, excitement, puzzlement, or "buzz" about the information set, similar to the measure of the decibels of noise in trading pits used by Coval and Shumway (2001). Activity measures are based on the notion that discussion (whether in person, by electronic posting, or news stories) is correlated with the salience and newness of information releases.

Capturing the content of the information is a more complicated matter. Although we did some of this by hand, this method is infeasible for a large data sample. Therefore, we extract a subjective measure of the meaning of the information by using computer algorithms that read and categorize the content of each individual message. The algorithms parse the degree to which the message conveys a buy, sell, or neutral sentiment about a stock. By aggregating these messages over some time period, we can gauge the average sentiment as well as the distribution of "posting sentiment" manifested by the stock message board information flow. We also use this method to classify the "news sentiment" of press stories. We call the combination of activity measures and content measures (distribution of sentiment indices) "eInformation."[3]

Sentiment is an intangible quality that is critical to many models used in financial economics.

[3]Tumarkin and Whitelaw (2001) study a subset of postings from one message board, which permits posters to voluntarily classify their short-term opinion about each stock. While these voluntary disclosures are convenient for study, only less than a quarter of posters choose to reveal a "short term opinion," and the board that permits this disclosure accounted for less than 5% of the total postings in our sample.

In behavioral finance, investor sentiment (or noise trader sentiment) is used to explain deviations in prices from "rational" levels (see Delong, Shleifer, Summers, and Waldman, 1990). To measure sentiment, academics have used the closed-end fund discount (Lee, Shleifer, and Thaler, 1991), flows into mutual funds (Goetzman, Massa, and Rouwenhorst, 2004), and subjective determinations based on reading various news stories (Gay, Kale, Kolb, and Noe, 1994). We create new measures of sentiment by examining posting or news stories.

Our measures of sentiment can be useful for a variety of reasons:

(a) Posting sentiment reflects the widely available opinions of a set of retail investors. The information is freely available on the Internet and the web sites that offer this information are widely visited.
(b) Newspaper and other media have always had the power to affect many people, but the Internet makes more news available to more potential investors more quickly. We have the capacity to "sign" the news to determine its likely impact on investor sentiment. Studies, such as those by Busse and Greene (2002) and Fleming and Remolona (1999), show rapid responses to alleged news events in the equity and bond markets.
(c) By using the disaggregated observation of sentiment, we can calculate various distribution measures. Not only can we calculate an average measure (net bulls less bears), but we can also calculate measures of the degree to which investors disagree, which is a variable that theoretical research suggests will correlate with both trading volumes and volatility (Kim and Verrecchia, 1991).
(d) The detailed nature of the data allows for stock-by-stock observation of sentiment that is more granular than broad market-wide sentiment indices. The high frequency of the data allows us to plot changes in sentiment over time.

B. Calculating eInformation Measures

The four firms in our study together received over 170,000 posts over seven months. However, our primary interest is to extract some meaning from the messages, in particular the "bullishness and bearishness" of the posts and the extent to which posters seem to agree or disagree.

We follow the Das and Chen (2003) method to classify the messages. We use a voting algorithm based on five underlying classifiers to improve the overall signal to noise ratio of the measure. Our approach results in a level of accuracy slightly lower than that of human classification.

We note that we developed the classification algorithms in this paper from several different ideas in the field of linear algebra and statistical theory. Earlier work in a different text classification domain comes from the work of Koller and Sahami (1997) and Chakrabarti, Dom, Agrawal, and Raghavan (1998).

The five classification routines use different rules to determine whether the message is a buy, a sell, or neutral. We then count the number of "votes" across the five different measures. We assign messages that receive at least three bullish (bearish or neutral) votes in that category; otherwise they are not categorized (nc). Although the unit of observation for classification is the message, we also create daily measures of the eInformation, as well as measures for other time intervals.

Our primary measure is a sentiment index, which we define as the number of buy messages less the number of sell messages (excluding null and not classified messages). The sentiment index picks up the net bullish sentiment and is an "absolute" or unscaled measure of sentiment.

In addition, we calculate a number of other related measures:

- Sentiment sign: one if sentiment > 0; -1 if sentiment <0, zero otherwise.
- Sentiment percentage: Sentiment index divided by all messages for the day. This is essentially a scaled measure of sentiment. It assesses the extent of the day's discussion that comprises bullish or bearish opinion. We include null messages as well as ambiguous messages in the total count in the denominator of this measure because they are both symptomatic of the absence of strong sentiment.
- Opinion index: Fraction of all messages that we classify as either buy or sell. One interpretation of the complement of the opinion index (i.e., 1-index) is that it represents the extent of questions, non-directional comments and noise in the discussion.
- Disagreement index: We define this index as:

$$\left| \frac{|BUYS - SELLS|}{BUYS + SELLS} - 1 \right| \qquad (1)$$

(or n.a. if Buys+Sells = 0).

This measure is intended to capture whether the opinionated posters have the same view, or whether there is dispersion of belief. If everyone is on the buy or sell side of the market, this index is zero, but the index can rise to 100% (or one) if the opinions are split equally into buys and sells. To the extent that total messages are highly correlated with signed messages, the disagreement index is like an unsigned version of the sentiment percentage.

Table IV reports the overall categorization of messages for the four stocks over the entire sample period. For the four firms, we can classify about 40% to 50% of the messages as either buys or sells. About 6% to 8% of the total posts are net buys. This low level of positive sentiment reflects the fact that there tend to be large numbers of both buy and sell messages, shown by the disagreement index of 80% or more (100% would mean that the buys and sells are equally split). This high disagreement level is not surprising, given that discussion takes place when there are differences of opinion, and given the fact that longs and shorts can both participate on the boards.

V. The Analysis of Sentiment and Disagreement

Investor opinion is likely shaped by a variety of forces. Here, we report on how several variables relate to our two measures of interest, sentiment and disagreement.

First, we expect that opinions change slowly, so that there will be persistence in the sentiment and disagreement time series, measurable by the levels of autocorrelation of these series.

Second, by discussing their preferences, posters have an obvious interest in stock returns and volatility. If the posters' impressions are formed by the level of prices, we should see a positive relation between returns and sentiment. However, if the posters tend to be contrarians, we could find a negative relation. If volatility represents uncertainty, then high levels of volatility might be associated with more disagreement, given the overall level of uncertainty. We measure returns with data from CRSP. We measure forward-looking uncertainty with implied volatilities on short-dated options reported on Bloomberg.

Third, we might see evidence of "reinforced persistence." Investor's views are likely to be more persistent when they are reinforced by data, suggesting that bullish sentiment is likely

Table IV. eInformation Variables for the Four Companies Studied

This table shows the classification of the messages posted on the four major stock message boards (Yahoo!, The Motley Fool, Silicon Investor, and Raging Bull) for the four stocks for the period July 1, 1998 through January 31, 1999. We define "Opinion" as the percentage of all messages that are either buys or sells. We define "Sentiment" as the net number of buy minus sell messages. "Sentiment %" divides the sentiment measure by the total number of messages. We define "Disagreement" as ||Sentiment|/(Buy + Sell Messages) – 1|.

	AMZN		DAL		GMGC		GWRX	
Total Period Messages								
Buy	29,367	29%	404	29%	15,276	24%	557	28%
Sell	23,017	22%	293	21%	10,949	17%	372	19%
Neutral	36,916	36%	535	38%	31,854	49%	813	41%
Nonclassified	13,363	13%	166	12%	6,835	11%	236	12%
Total	102,663	100%	1,398	100%	64,914	100%	1,978	100%
Opinion	51%		50%		40%		47%	
Sentiment	6,350		111		4,327		185	
Sentiment %	6%		8%		7%		9%	
Disagreement %	88%		84%		84%		80%	
Daily Averages								
No message days	0		10		0		20	
Opinion								
Mean	52%		45%		41%		45%	
Median	52%		50%		41%		50%	
Std Deviation	5%		25%		5%		27%	
Sentiment								
Mean	30		1		20		1	
Median	20		0		17		0	
Std Deviation	43		2		17		3	
Disagreement								
Mean	85%		43%		80%		43%	
Median	87%		50%		82%		50%	
Std Deviation	10%		41%		13%		40%	

to be more autocorrelated when returns are positive.

Fourth, we suspect that people turn from "lurkers" to "posters" either when they have questions or strong opinions. Changes in the level of sentiment and increasing levels of disagreement might be related to the level of "new posters" in the group.

Fifth, discussion is more likely when various parties disagree or when the level of sentiment is high. Therefore, we anticipate that the level of posting activity (controlling for day of the week effects, which are meaningful for posting activity) will be related to the levels of sentiment and disagreement, with higher posting activity related to higher absolute levels of sentiment (either positive or negative) and greater disagreement.

Sixth, our discussion with Glenn and our analysis of the content of postings suggests that these on-line discussions take place in an environment in which posters are collecting and studying news, filings, analyst reports, and nonfinancial data. We expect to find a positive relation between news sentiment (extracted from news stories using the same algorithm) and posting sentiment, and between news disagreement and posting disagreement.

To measure news sentiment and disagreement, we apply the algorithm described above to the major news stories on Factiva. As before, we define a "major news story" as one in which the company's name is either in the headline or mentioned in the lead and at least three times in the body of the article. In addition, we collected press releases and filing information from Factiva, Edgar, and Global Access, as well as analyst reports and earnings revisions from Investext and from IBES.

Our sample includes information that can be easily obtained by a retail investor without real-time monitoring, but excludes TV and radio broadcasts, which were not available to us. We also exclude information that might be available only to large institutional investors (e.g., conference call proceedings prior to web broadcasting or private communications with management prior to Regulation FD. See Bushee, Matsumoto, and Miller, 2003).

Table V categorizes the information events for each of the four companies. Over the seven-month period, there are 168 press releases, 58 filings, 207 analyst forecast revisions, 1,667 major news stories, and 170,953 stock chat posts. The dispersion in the information releases is intentionally large, because we wish to capture four different types of firms. For example, there are 73 times as many postings at Amazon than at Delta, but 15 times more stories about Delta than about either General Magic or Geoworks.

Table VI reports the univariate statistics and definitions of the variables we use in our analyses in the remainder of the article.

A. Empirical Evidence on the Determinants of Sentiment and Disagreement

One of our goals is to understand what drives investor sentiment. To accomplish this goal we regress sentiment and disagreement measures on explanatory variables such as lagged values of sentiment, the current and lagged values of stock returns, posting volume, trading volume, lagged market return, and current and lagged sentiment derived from news sources using our algorithm.

In Table VII, Panel A shows the determinants of posting volume. There are more messages posted when the stock's trading volume is high and when there are new posters in the prior seven days. There is also evidence of significant persistence in posting volume for DAL and GMGC (but not for AMZN and GWRX). For AMZN and GMGC, there is a positive relation between the number of news stories and the level of message posting. As noted earlier, more active boards may use discussion to interpret and digest news releases, and occasionally, the news media reports on investor interest reflected on the message boards.

For all four stocks, but significantly only for GWRX and GMGC, there is a negative relation between contemporaneous returns and postings, suggesting that message volume picks up when the stocks do poorly. This result may be consistent with loss aversion—perhaps losses are more salient to message posters than gains. However, lagged returns are inconsistently related to volume.

We also find that disagreement is related to message volume. For all stocks but AMZN, contemporaneous disagreement is significant, though lagged disagreement is not (except for DAL, which is negative). Disagreement and discussion go hand in hand. The high adjusted R-squares and significant F-statistics suggest that we are able to model posting volume reasonably well.

Panel B suggests a high level of persistence in the sentiment level, as predicted. This persistence is shown by the significance of lagged sentiment in three of the four regressions. In addition, sentiment is also positively related to the total volume of postings across all four boards, to trading volume (although only significant for GWRX), and to current stock

Table V. Information Events for the Firms in the Sample

The table below shows the total number of information events (press releases, filings, analyst reports/revisions, major news stories, and posts) for the four sample firms over the time period July 1, 1998 through January 31, 1999.

	Amazon	Delta	General Magic	Geoworks	Total
News quintile	High	High	Low	Low	
Chat quintile	High	Low	High	Low	
Press releases	22	109	20	17	168
Filings	26	10	12	10	58
Analyst reports and revisions	135	68	—	4	207
Major news stories	987	549	66	65	1,667
Postings	102,663	1,398	64,914	1,978	170,953
Total	103,833	2,134	65,012	2,074	173,053

returns. These findings imply that small investors are attentive to market activity and influenced by it, and are consistent with other studies that find the same result (see Das and Chen, 2003 and Antweiler and Frank, 2004).

While we normally study how information (or sentiment) is impounded into stock prices, our analysis suggests how stock returns are impounded into sentiment. Table VII, Panel B, shows that not only is sentiment related to the contemporaneous return, but the prior day return as well, at least for the two stocks with active posting levels, AMZN and GMGC.[4] Except for GMGC, overall market returns do not influence stock-specific sentiment.

We test whether sentiment is more persistent when the market return reinforces the previous day's sentiment. In unreported regressions, we add a dummy variable equal to one if the lagged stock return and lagged posting sentiment had the same sign (and zero otherwise). This variable is not significant, and the results on other variables are not noticeably affected. Thus, we do not observe that sentiment persistence is affected by its recent accuracy.

In univariate results, reported in Table VIII, investor sentiment and news sentiment have strong positive correlations for all firms but GMGC. However, in this multivariate setting, news sentiment is unrelated to message board sentiment (with the curious exception of GMGC). It is likely that other more fundamental factors, i.e., current returns, are more fundamental drivers, so that news per se adds limited incremental content.

In Panel C, we see that disagreement is primarily related to message posting volume. As expected, disagreement and discussion go hand-in-hand. Disagreement is significantly persistent for AMZN and GMGC, which have high posting volumes.

Surprisingly, few of the other variables are correlated with disagreement. Implied volatility is unrelated to disagreement, even though we thought it might capture the future uncertainty of returns for a stock. From the intercept term, we can get some idea of the range of base-level disagreement, which implies that the difference between bullish and bearish messages is about 35% to 70% of total signed messages.

Overall, we find that sentiment changes slowly and that it is related to stock returns and trading volume. Higher message volume and disagreement are related to one another. However, our daily evidence does not suggest that disagreement leads to discussion, as we had anticipated.

[4]We are grateful to an anonymous referee for suggesting this analysis.

Table VI. Variable Definitions

Variable Definition	Summary Statistics: Mean (Std. Deviation) All Firm-Periods	AMZN	DAL	GMGC	GWRX
Implied Volatility: Implied volatility on at-the-money call options from Bloomberg.	0.880 (0.354)	0.948 (0.168)	0.450 (0.074)	1.243 (0.140)	—
Share Turnover: Number of shares traded that day divided by number of shares outstanding.	0.040 (0.054)	0.096 (0.051)	0.009 (0.004)	0.036 (0.038)	0.020 (0.057)
Announcements					
Press Release: Dummy variable. Equals one if the company has made a press release on that day.	0.209 (0.374)	0.142 (0.350)	0.459 (0.500)	0.135 (0.343)	0.101 (0.303)
Filing: Dummy variable. Equals one if the company made an SEC filing that day.	0.090 (0.281)	0.149 (0.357)	0.061 (0.240)	0.081 (0.274)	0.068 (0.252)
Analyst Revision: Dummy variable. Equals 1 if any analyst issued some sort of earnings revision that day.	0.122 (0.327)	0.250 (0.434)	0.216 (0.413)	—	0.020 (0.141)
News Activity					
Abnormal News Stories: Residual from regression of news stories on prior day news stories, day of week, and month of year dummy variables.	0.000 (2.996)	0.000 (4.765)	0.000 (5.445)	0.000 (0.822)	0.000 (0.951)
Lagged Abnormal Stories: Note: lag determined by last calendar day (Monday lag includes weekend)					
Posting Activity					
Abnormal Number of Posts (Close-to-Close): Residual from regression of posts on the 4 p.m. prior day to 4 p.m. trading day period, on its lag, and day of week, and month of year dummies.	0.000 (105.9)	0.000 (181)	0.000 (2.43)	0.000 (109.2)	0.000 (14.42)
Abnormal Number of Market Posts: Residual from regression of posts on trading day from 9:30 a.m. to 4 p.m. on its lag, and day of week, and month of year dummy variables.	0.000 (52.20)	0.000 (82.28)	0.000 (2.29)	0.000 (63.86)	0.000 (9.26)
Abnormal Number of Pre-Market Posts: Residual from regression of posts from 4 p.m. prior day to 9:30 a.m. on trading day on its lag, and day of week, and month of year dummy variables.	0.000 (77.16)	0.000 (133.7)	0.000 (3.03)	0.000 (76.96)	0.000 (9.08)
Sentiment Level (Close-to-Close): Number of buy messages – number of sell messages from 4:00 p.m. prior day to 4:00 p.m. trading day.	12.682 (24.42)	30.723 (48.42)	0.581 (1.96)	23.034 (18.20)	0.973 (3.50)
Sentiment Level during Market Hours: Number of buy messages – number of sell messages from 9:30 a.m. to 4:00 p.m.	4.973 (13.46)	9.378 (23.25)	0.284 (1.19)	9.655 (10.01)	0.574 (1.88)
Sentiment Level during Pre-Market Hours: Number of buy messages – number of sell messages from 4:00 p.m. prior trading day to 9:30 a.m. on the focal trading day.	8.855 (19.2)	21.345 (31.22)	0.297 (1.46)	13.378 (13.22)	0.399 (2.86)
Interaction Terms: Note: the information events are dummy variables indicating a press release (etc.) on the current day. The four interaction terms indicate the abnormal level of news stories or posts, or the level of news sentiment or posting sentiment.					
Lags: Note: lags are determined by previous trading day, not by previous calendar day.					

Table VII. Analysis of Posting Activity Characteristics

In this table the dependent variables are the different dimensions of eInformation: Posting Volume, Sentiment, and Disagreement. We pool the data from the four boards for each of the four firms during the period 7/1/98-1/31/99.

Panel A. Analysis of Posting Volume

	AMZN		DAL		GMGC		GWRX	
Dependent Variable: Number of Postings	Coeff	p-value	Coeff	p-value	Coeff	p-value	Coeff	p-value
Constant	-112.390	(0.551)	3.006	(0.456)	-75.038	(0.531)	-7.213	(0.022)
Posting Activity								
Lagged number of postings	0.015	(0.827)	0.319	(0.001)	0.443	(0.000)	0.059	(0.372)
Disagreement	48.375	(0.774)	4.645	(0.000)	172.958	(0.035)	8.777	(0.005)
Lagged disagreement	87.219	(0.610)	-2.742	(0.018)	38.120	(0.631)	3.313	(0.286)
Stock Market Activity								
Trading volume	0.000	(0.000)	0.000	(0.147)	0.000	(0.000)	0.000	(0.000)
Stock return	-276.077	(0.159)	-18.428	(0.246)	-448.101	(0.001)	-28.416	(0.017)
Lagged stock return	-293.856	(0.092)	31.059	(0.031)	-136.846	(0.229)	30.783	(0.000)
Equal weighted stock market return	339.337	(0.756)	-43.126	(0.319)	712.895	(0.332)	141.738	(0.239)
Implied volatility	37.369	(0.687)	-4.232	(0.562)	-85.237	(0.188)	—	—
News Activity								
Average news sentiment	-8.835	(0.133)	0.389	(0.423)	-2.171	(0.935)	-3.957	(0.356)
Abnormal news stories	6.466	(0.004)	0.137	(0.337)	26.168	(0.004)	-2.078	(0.158)
New posters in the last seven days	0.593	(0.000)	0.022	(0.817)	0.625	(0.000)	0.252	(0.000)
Number of observations		143		114		144		100
Adjusted R-squared		0.870		0.311		0.772		0.870
F-statistic		56.51		4.40		33.22		48.23

Table VII. Analysis of Posting Activity Characteristics (*Continued*)

Panel B. Analysis of Sentiment

	AMZN		DAL		GMGC		GWRX	
Dependent variable: Sentiment level	**Coeff**	**p-value**	**Coeff**	**p-value**	**Coeff**	**p-value**	**Coeff**	**p-value**
Constant	-30.262	(0.040)	-0.468	(0.745)	-5.927	(0.633)	-0.550	(0.209)
Posting Activity								
Lagged sentiment	0.271	(0.002)	-0.087	(0.337)	0.134	(0.080)	0.189	(0.017)
Number of postings	0.034	(0.035)	0.125	(0.000)	0.034	(0.004)	0.061	(0.001)
Stock Market Activity								
Trading volume	0.000	(0.928)	0.000	(0.203)	0.000	(0.190)	0.000	(0.076)
Stock return	104.340	(0.004)	12.635	(0.040)	41.548	(0.045)	-0.734	(0.689)
Lagged stock return	122.188	(0.000)	5.438	(0.319)	58.101	(0.001)	1.660	(0.282)
Equal weighted stock market return	208.854	(0.294)	15.257	(0.331)	195.921	(0.074)	-21.798	(0.165)
Implied volatility	13.120	(0.432)	0.073	(0.979)	2.323	(0.813)	—	—
News Activity								
Average news sentiment	1.387	(0.194)	-0.043	(0.817)	6.908	(0.095)	0.239	(0.726)
Abnormal news stories	0.285	(0.483)	0.065	(0.252)	0.241	(0.862)	-0.371	(0.139)
New posters in the last seven days	0.034	(0.051)	0.017	(0.641)	0.030	(0.220)	0.014	(0.247)
Number of observations	143		144		144		144	
Adjusted R-squared	0.733		0.096		0.425		0.645	
F-statistic	28.890		2.080		8.550		20.950	

Table VII. Analysis of Posting Activity Characteristics (*Continued*)

Panel C. Analysis of Disagreement

	AMZN		DAL		GMGC		GWRX	
Dependent Variable: Disagreement in posting activity	**Coeff**	**p-value**	**Coeff**	**p-value**	**Coeff**	**p-value**	**Coeff**	**p-value**
Constant	0.560	(0.000)	0.434	(0.211)	0.716	(0.000)	0.341	(0.001)
Posting Activity								
Lagged disagreement	0.377	(0.000)	0.106	(0.295)	0.169	(0.046)	0.086	(0.416)
Number of postings	0.000	(0.774)	0.035	(0.000)	0.000	(0.035)	0.010	(0.005)
Lagged number of postings	0.000	(0.819)	0.000	(0.980)	0.000	(0.615)	-0.001	(0.665)
Stock Market Activity								
Trading volume	0.000	(0.682)	0.000	(0.709)	0.000	(0.029)	0.000	(0.128)
Stock return	-0.125	(0.228)	0.016	(0.991)	-0.205	(0.157)	0.363	(0.379)
Lagged stock return	-0.090	(0.328)	-2.205	(0.078)	-0.230	(0.058)	-0.464	(0.132)
Equal-weighted stock market return	0.138	(0.811)	1.835	(0.624)	-1.470	(0.061)	-2.083	(0.613)
Implied volatility	0.001	(0.984)	-0.445	(0.480)	-0.026	(0.704)	—	—
News Activity								
Average news sentiment	-0.003	(0.271)	-0.035	(0.410)	-0.033	(0.245)	0.098	(0.504)
Abnormal news stories	-0.001	(0.482)	-0.007	(0.575)	0.003	(0.774)	0.083	(0.096)
New posters in the last 7 days	0.000	(0.399)	0.002	(0.822)	0.000	(0.543)	-0.001	(0.647)
Number of observations	143		114		144		100	
Adjusted R-squared	0.262		0.13		0.228		0.057	
F-statistic	4.35		2.13		3.81		143	

VI. eInformation and the Price-Formation Process

Market efficiency implies that public information is immediately embedded in the stock price (Fama, 1965 and 1975). Given our discussions with Glenn and our inspection of news stories, we were skeptical whether the message boards could predict future returns. Indeed, they do not, as we show below.

A. Correlations

In Table VIII we report the contemporaneous correlations between our information variables and market variables (returns, excess returns, share turnover, implied volatility and intraday volatility, and average bid-ask spreads) as well as the autocorrelations of market variables.

For the two small firms, and to a lesser extent for Amazon, there is a significant contemporaneous positive correlation between the number of news stories and stock returns. This result is consistent with the press writing about high performing stocks. For General Magic and Geoworks, there is contemporaneous positive correlation between news sentiment and returns, which indicates favorable media reports about high performing stocks. For all but the low-news/low-posting Geoworks), there is significant contemporaneous correlation between message board sentiment measures and returns. For the two firms with substantial chat activity, there is a negative correlation between disagreement and returns. When people disagree, returns tend to be lower. Or conversely, when stocks fall, there tends to be more discussion and greater disagreement.

While the contemporaneous correlation between information variables and returns is modest, the contemporaneous correlation between the eInformation variables and turnover, volatility, bid-ask spreads and jumps is more robust, especially for the most actively discussed firms in our sample, Amazon and General Magic. These non-return aspects of the financial markets are also correlated with one another, as shown in Table VIII.

B. Are Returns for these Four Stocks Explicable Using eInformation?

For the four stocks we study, we examine whether the eInformation and other information variables help to explain returns. If the eInformation variables are meaningful, at a minimum we should observe contemporaneous correlations between them and their returns. If they contain truly new information, we might observe that eInformation predicts returns.

We examine not only close-to-close returns, but also open-to-close returns. This latter measure tests whether reading posts from the day before and the night/early morning prior to market opening would permit a trader to predict subsequent returns. We use a specification that is similar to those used by Wysocki (1999) and Mitchell and Mulherin (1994). Table VI defines our variables.

We include both contemporaneous and lagged information variables. Including contemporaneous variables tests if our measures are sheer noise or whether they are picking up signals that confirm the current state of the market. Lagged information sources help determine whether the information can predict returns.

Table IX provides the results of this inquiry. Panel A looks at the relation between contemporaneous information measures and close-to-close market-adjusted returns and Panel B uses open-to-close market adjusted returns. Both of these panels show whether returns are related to the same-day announcements, news activity, and posting activity. Evidence of significant relations would not suggest market inefficiency, but would instead be consistent

Table VIII. Correlations between Information and Financial Markets

This table reports the same-day contemporaneous correlations between information variables and financial market variables (return, excess return, volume, implied volatility, intraday volatility, and bid-ask spreads). Each cell represents the median correlation among the four firms, the median p-value among the four firms, and the identity of the firms for which the p-value is 0.05 or better. In each cell, a=Amazon, d=Delta, m= General Magic and x = Geoworks.

	Press Release	Filing	Analyst Revision	# Major News Stories	News Sentiment	Number of Postings	New Posters Last 7days	Analyst Sentiment
Filing	0.024 0.768	1.000						
Analyst Revision	-0.043 0.352	-0.032 0.660 a	1.000					
Number of Major News Stories	0.601 0.000 admx	-0.019 0.821	0.072 0.383 a	1.000				
News Sentiment	0.327 0.000 dmx	-0.004 0.779	0.013 0.879	0.390 0.000 admx	1.000			
Number of Postings	0.058 0.485 m	-0.013 0.519	0.150 0.070 a	0.284 0.010 amx	-0.021 0.209 ax	1.000		
New Posters in Last 7 Days	0.040 0.647 m	-0.095 0.302	0.111 0.179 x	0.222 0.047 ax	0.010 0.384 a	0.728 0.000 admx	1.000	
Analyst Sentiment	-0.089 0.284 a	0.000 0.437 a	-0.518 0.000 ad	-0.101 0.224 a	-0.012 0.890 a	0.108 0.067 a	-0.128 0.012 ax	1.000
Sentiment	0.055 0.571 m	0.014 0.364	0.026 0.753	0.251 0.006 amx	0.096 0.014 adx	0.562 0.000 admx	0.500 0.000 amx	0.080 0.069 a

Table VIII. Correlations between Information and Financial Markets (*Continued*)

	Press Release	Filing	Analyst Revision	# Major News Stories	News Sentiment	Number of Postings	New Posters Last 7days	Analyst Sentiment
Disagreement	0.023	0.019	0.100	-0.075	-0.107	0.269	0.111	-0.028
	0.785	0.367	0.254	0.248	0.252	0.001	0.047	0.110
				a	a	admx	ax	
Open-to-Close Return	0.030	-0.010	-0.040	0.176	0.147	-0.044	-0.002	0.119
	0.520	0.563	0.631	0.054	0.296	0.601	0.652	0.152
	x			mx	mx			
Close-to-Close Return	0.088	-0.018	0.022	0.248	0.265	-0.066	0.038	0.029
	0.375	0.727	0.793	0.012	0.023	0.521	0.623	0.705
	mx			amx	amx			
Equal Wtd Market Return	-0.026	0.017	0.090	-0.031	-0.028	-0.040	0.051	0.030
	0.726	0.760	0.275	0.708	0.745	0.383	0.409	0.715
			x			m		a
Share Turnover	0.196	0.035	0.001	0.407	0.209	0.495	0.336	-0.028
	0.041	0.446	0.988	0.000	0.152	0.000	0.001	0.534
	mx	m		admx	am	amx	amx	
Intraday Volatility	0.028	0.008	0.073	0.171	0.119	0.120	-0.051	-0.073
	0.737	0.561	0.457	0.040	0.249	0.186	0.175	0.380
				am	am	am	ad	
Implied Volatility	0.017	-0.023	0.069	0.049	-0.023	0.030	-0.164	-0.215
	0.080	0.785	0.451	0.555	0.633	0.227	0.000	0.073
				a		a	adm	a
Bid-Ask %	-0.088	-0.094	-0.014	-0.259	-0.063	-0.367	-0.395	-0.010
	0.181	0.258	0.865	0.002	0.435	0.000	0.000	0.902
	m	x		amx	a	admx	amx	
Number of Jumps	0.085	-0.014	0.071	0.136	0.128	0.166	0.206	-0.154
	0.311	0.533	0.389	0.113	0.181	0.083	0.015	0.062
	x			x	am	mx	amx	

Table VIII. Correlations between Information and Financial Markets (*Continued*)

	Sentiment	Dis-agreement	Open-to-Close Return	Close-to-Close Return	Equal Wtd. Mkt. Return	Share Turnover	Intraday Volatility	Implied Volatility	Bid-Ask %
Disagreement	-0.341	1.000							
	0.120								
	am								
Open-to-Close Return	0.103	-0.048	1.000						
	0.217	0.577							
	d	m							
Close-to-Close Return	0.210	-0.093	0.902	1.000					
	0.011	0.340	0.000						
	amd	am	admx						
Equal Wtd. Mkt. Return	0.133	-0.070	0.243	0.371	1.000				
	0.146	0.410	0.003	0.000					
	ad	m	adm	adm					
Share Turnover	0.321	0.018	0.184	0.217	-0.043	1.000			
	0.000	0.172	0.050	0.017	0.615				
	amx	a	ax	amx					
Intraday Volatility	0.028	-0.010	0.127	0.089	-0.169	0.436	1.000		
	0.523	0.597	0.207	0.114	0.116	0.000			
	a		m		am	adm			
Implied Volatility	-0.031	-0.060	-0.151	-0.190	-0.024	0.185	0.450	1.000	
	0.332	0.495	0.069	0.022	0.769	0.025	0.000		
	a		m	am	m	ad	adm		
Bid-Ask %	-0.287	0.056	-0.101	-0.149	-0.038	-0.276	0.360	0.257	1.000
	0.001	0.168	0.226	0.079	0.633	0.001	0.001	0.002	
	amx	ax	a	am	a	amx	admx	dm	
Number of Jumps	0.145	0.015	0.042	0.072	0.047	0.186	-0.103	0.057	-0.185
	0.338	0.386	0.404	0.394	0.296	0.031	0.073	0.214	0.028
	ax	x	x	x		mx	dm		admx

Table IX. Explaining Returns Using Information Variables

In this table the dependent variables are the market-adjusted return (raw return less S&P 500) earned by each of the four firms during the period 7/1/98–1/31/99. For Panels A and C, we calculate this return from the close of the market on the prior trading day to the close of the market on the current day. For Panels B and D, the return represents the return from the open of the market to the close of the market. Panels A and B use contemporaneous information, i.e., information from the same day as the return. This information includes all filings, press releases, and news stories with that date, and postings for the same period for which we calculate returns. Panels C and D use lagged information, i.e., filings, press releases, and news stories from the prior day, and postings that preceded the return computation period.

Panel A.

Dependent variable: market adjusted returns close-to-close	AMZN		DAL		GMGC		GWRX	
All independent variables measured the same "day."	**Coeff**	**p-value**	**Coeff**	**p-value**	**Coeff**	**p-value**	**Coeff**	**p-value**
Constant	0.142	(0.000)	0.023	(0.066)	0.175	(0.000)	0.009	(0.646)
Volatility	-0.152	(0.000)	-0.054	(0.043)	-0.153	(0.000)	0.000	(0.991)
Announcements								
Press Release	0.037	(0.136)	-0.001	(0.891)	-0.115	(0.000)	-0.342	(0.000)
Filing	-0.002	(0.877)	-0.011	(0.110)	0.003	(0.890)	-0.012	(0.733)
Revision	-0.002	(0.815)	0.001	(0.899)	—	—	0.004	(0.856)
News Activity								
Abnormal number of stories	0.002	(0.113)	0.001	(0.548)	0.028	(0.008)	0.016	(0.381)
News sentiment	-0.001	(0.537)	-0.002	(0.576)	0.020	(0.076)	0.056	(0.006)
Posting Activity								
Abnormal number of market posts	-0.000	(0.934)	-0.001	(0.365)	-0.000	(0.067)	0.000	(0.889)
Sentiment level	0.000	(0.006)	0.002	(0.174)	0.000	(0.071)	-0.004	(0.249)
Interaction terms: Press release dummy interacted with…								
…abnormal number of stories	-0.000	(0.906)	-0.002	(0.447)	0.018	(0.299)	0.142	(0.000)
…sentiment of stories	-0.016	(0.007)	0.003	(0.392)	0.017	(0.249)	-0.030	(0.398)
…abnormal number of market posts	0.000	(0.145)	-0.001	(0.474)	-0.000	(0.336)	0.008	(0.194)
…sentiment of posts	0.001	(0.008)	0.001	(0.581)	0.001	(0.061)	0.022	(0.087)
Number of observations	146		147		147		105	
Adjusted R-squared	0.189		0.101		0.443		0.754	
F-statistic	3.42		1.82		8.23		16.13	

Table IX. Explaining Returns Using Information Variables (*Continued*)

Panel B.

Dependent variable: market adjusted returns, open-to-close All independent variables measured the same "day."	AMZN		DAL		GMGC		GWRX	
	Coeff	**p-value**	**Coeff**	**p-value**	**Coeff**	**p-value**	**Coeff**	**p-value**
Constant	0.111	(0.000)	0.008	(0.576)	0.158	(0.000)	-0.009	(0.660)
Volatility	-0.107	(0.002)	-0.018	(0.530)	-0.143	(0.000)	0.004	(0.598)
Announcements								
Press Release	0.007	(0.710)	0.000	(0.961)	-0.101	(0.000)	-0.326	(0.000)
Filing	-0.010	(0.405)	-0.011	(0.093)	0.000	(0.997)	0.066	(0.065)
Revision	-0.008	(0.403)	-0.004	(0.478)	—	—	0.026	(0.575)
News Activity								
Abnormal number of stories	0.003	(0.025)	0.001	(0.543)	0.023	(0.015)	0.025	(0.196)
News sentiment	-0.003	(0.030)	-0.001	(0.738)	0.006	(0.591)	0.052	(0.001)
Posting Activity								
Abnormal number of market posts	0.000	(0.238)	0.001	(0.510)	0.000	(0.375)	-0.001	(0.043)
Sentiment level	0.001	(0.001)	0.003	(0.403)	0.001	(0.021)	-0.007	(0.326)
Interaction terms: Press release dummy interacted with…								
…abnormal number of stories	-0.001	(0.694)	-0.002	(0.407)	0.010	(0.472)	0.113	(0.002)
…sentiment of stories	-0.003	(0.484)	0.002	(0.566)	0.018	(0.184)	0.004	(0.855)
…abnormal number of market posts	0.001	(0.057)	-0.002	(0.175)	0.000	(0.767)	0.005	(0.386)
…sentiment of posts	0.001	(0.188)	0.001	(0.785)	0.002	(0.188)	0.039	(0.011)
Number of observations	146		147		147		105	
Adjusted R-squared	0.229		0.070		0.371		0.745	
F-statistic	7.02		1.12		9.85		23.14	

Table IX. Explaining Returns Using Information Variables (*Continued*)

Panel C.

Dependent variable: market adjusted returns close-to-close All independent variables measured the prior "day".	AMZN		DAL		GMGC		GWRX	
	Coeff	**p-value**	**Coeff**	**p-value**	**Coeff**	**p-value**	**Coeff**	**p-value**
Constant	0.048	(0.695)	-0.010	(0.480)	-0.057	(0.240)	-0.020	(0.555)
Volatility	-0.043	(0.215)	0.021	(0.454)	0.047	(0.229)	0.017	(0.319)
Announcements								
Press Release	0.005	(0.844)	0.002	(0.791)	-0.028	(0.355)	0.042	(0.342)
Filing	-0.010	(0.510)	0.001	(0.923)	-0.010	(0.548)	-0.064	(0.377)
Revision	-0.009	(0.558)	-0.001	(0.837)	—	—	0.035	(0.164)
News Activity								
Abnormal number of stories	0.000	(0.723)	0.000	(0.965)	0.006	(0.674)	-0.010	(0.685)
News sentiment	0.000	(0.914)	-0.001	(0.852)	0.008	(0.636)	0.016	(0.238)
Posting Activity								
Abnormal number of posts	0.000	(0.450)	-0.001	(0.169)	0.000	(0.813)	-0.001	(0.196)
Sentiment level	0.000	(0.550)	0.002	(0.208)	0.000	(0.550)	-0.003	(0.367)
Interaction terms: Press release dummy interacted with…								
…abnormal number of stories	-0.001	(0.724)	-0.001	(0.728)	-0.020	(0.224)	-0.048	(0.065)
…sentiment of stories	0.003	(0.781)	-0.002	(0.602)	0.000	(0.986)	0.039	(0.092)
…abnormal number of posts	0.000	(0.981)	0.001	(0.252)	0.000	(0.326)	0.005	(0.105)
…sentiment of posts	0.000	(0.142)	-0.004	(0.142)	0.000	(0.901)	-0.008	(0.438)
Number of observations	145		146		146		104	
Adjusted R-squared	0.050		0.072		0.118		0.075	
F-statistic	0.890		1.200		2.130		17.210	

Table IX. Explaining Returns Using Information Variables (*Continued*)

Panel D.

Dependent variable: market adjusted returns, open-to-close All independent variables measured the prior "day".	AMZN		DAL		GMGC		GWRX	
	Coeff	**p-value**	**Coeff**	**p-value**	**Coeff**	**p-value**	**Coeff**	**p-value**
Constant	0.036	(0.366)	-0.019	(0.164)	-0.078	(0.064)	-0.029	(0.342)
Volatility	-0.027	(0.570)	0.041	(0.120)	0.058	(0.071)	0.017	(0.300)
Announcements								
Press Release	0.005	(0.791)	-0.001	(0.808)	0.003	(0.898)	0.020	(0.686)
Filing	-0.019	(0.167)	0.002	(0.756)	-0.010	(0.477)	-0.106	(0.108)
Revision	-0.012	(0.359)	-0.002	(0.710)	—	—	-0.019	(0.488)
News Activity								
Abnormal number of stories	0.001	(0.596)	-0.001	(0.566)	0.000	(0.966)	-0.020	(0.388)
News sentiment	0.001	(0.634)	0.000	(0.975)	0.008	(0.633)	0.041	(0.008)
Posting Activity								
Abnormal number of pre-market posts	0.000	(0.479)	0.001	(0.465)	0.000	(0.920)	-0.003	(0.571)
Sentiment level	0.000	(0.967)	0.002	(0.367)	0.000	(0.889)	-0.005	(0.349)
Interaction terms: Press release dummy interacted with…								
…abnormal number of stories	0.003	(0.354)	0.000	(0.814)	-0.025	(0.045)	-0.014	(0.735)
…sentiment of stories of pre-market posts	-0.001	(0.849)	-0.002	(0.431)	-0.009	(0.608)	-0.003	(0.929)
…abnormal number of pre-market posts	0.000	(0.664)	-0.001	(0.378)	0.000	(0.400)	0.003	(0.642)
…sentiment of pre-market posts	-0.001	(0.224)	0.002	(0.563)	0.000	(0.995)	0.000	(0.981)
Number of observations	145		146		146		104	
Adjusted R-squared	0.069		0.082		0.118		0.083	
F-statistic	1.950		1.290		4.170		37.840	

with information production being impounded quickly into prices. Panels C and D repeat the analysis in the first two panels, except that all of the independent variables are lagged. Announcements and news levels are from the prior day, postings are from the pre-market period (4:00 p.m. prior day to 9:30 a.m. trading day). Relationships here suggest that a trader could observe data today (until the open of the market) and profit from it.

In Panels A and B, which examine contemporaneous information flows and returns, we see indications that stock returns may be higher on days when:

- There are more news stories (or vice versa). This relationship is statistically significant for the two high eInformation boards (AMZN, GMGC), implying a possible catalyst relationship from message board activity, when eInformation crosses a threshold level.
- This news conveys more positive sentiment, as measured by our sentiment algorithm applied to the text of news stories (GMGC, GWRX). Here, the effect is statistically significant for the two boards that have low news volume. The observation that returns and news are more related for low-news stocks suggests a certain diminishing marginal impact of news stories.
- The message board postings reflect greater positive sentiment (especially for those stocks with more active postings, AMZN and GMGC). This finding is consistent with those of Antweiler and Frank (2002).
- There is a press release in a context with more positive posting sentiment (AMZN, GMGC, GWRX). Separately, stock returns seem to be lower on days when low traditional information firms issue press releases (GMGC, GWRX), but this effect is mitigated when the company press release is combined with more news (especially GWRX).

Our posting sentiment index is more closely related to contemporaneous prices than is the sheer numbers of postings, which is the measure used by Wysocki (2000). These results are encouraging, in that they suggest that our sentiment index, applied to either short messages or longer news stories, captures the tone of the text. Furthermore, news and sentiment measures can help us to “sign” various news events, in this case, company press releases. Not surprisingly, press releases seem to gain importance as they are interpreted by the news and by board posters.

The results in Table IX, Panels C and D provide no evidence to support the idea that postings predict returns. Using information that arrives before the opening of the market, including overnight posting activity, we find that no “information” variable is consistently informative to a trader who will transact over the course of the day. Of the 43 firm-coefficients in each panel (four stocks times ten or 11 variables), three or four are significant at the 10% level, just as chance would predict.

We perform a sensitivity check on all these analyses using changes, as opposed to levels, of sentiment. We define changes as differences in sentiment, percentage differences in sentiment, and residuals from a regression on lagged sentiment and calendar variables. The results are directionally similar.

Our results are consistent with those found by Tumarkin and Whitelaw (2001) and by Antweiler and Frank (2004). All three papers were independently produced and use different samples and different methods for coding the information content of the message boards. Nevertheless, all three papers show no predictive power for the message boards that explains subsequent stock returns. In a word, all three of these small sample papers suggest that people trade first and talk later, with returns preceding postings, rather than the other way around.

VII. Discussion and Summary

We perform a clinical study of the process of investor discussion and sentiment formation, using stock message boards as a window into investor behavior. Using various language processing routines, we create sentiment and disagreement measures based on the comments posted on the message boards.

We find that a small core of members of online communities carry out an extended discussion that has several positive attributes. Our investigation suggests a set of hypotheses, which are that the boards provide readers with a community of like-minded investors, and that the group delivers on-point discussions, quick dissemination of new public information, and in some instances provides foreshadowing of subsequent news releases. However, these benefits come at the cost of a large number of "false positives" in the form of unsubstantiated rumors.

Further, we hypothesize that the calculus that leads someone to spend a lot of time investing in these discussions has more to do with sharing opinions than sharing (or collecting) any private information. We posit this hypothesis on the basis of our interview with an extensive poster, from our inspection of the specific content of the boards, and from our econometric analysis of the predictability of returns using sentiment. Although some posters apparently value the benefits of testing their ideas with others, a cost may be to harden the opinions of some.

We extract a time series of sentiment and disagreement measures from the message boards. We find that there is a close relation between sentiment levels and lagged sentiment, posting activity, stock returns, and lagged stock returns, but not news sentiment. In addition, disagreement is related to the intensity of discussion. The discussion on the boards provides one mechanism by which sentiment is created and firmed up. Thus, we provide an in-depth study of the mechanism driving investor sentiment.

Sentiment does not apparently predict returns, but returns drive sentiment. This finding suggests that members of the on-line community are more likely to extrapolate past returns, rather than to be contrarian, which leads to behavior consistent with the representativeness heuristic (Lakonishok, Shleifer, and Vishny, 1994).

Our extensive documentation of the environment on the message boards is also consistent with the idea that people will engage in social interaction to mitigate the costs of bounded rationality, and for opportunistic reasons (Baker, 1984).■

References

Antweiler, W. and M. Frank, 2002, "Internet Stock Message Boards and Stock Returns," University of British Columbia Working Paper.

Antweiler, W. and M. Frank, 2004, "Is All That Talk Just Noise? The Information Content of InternetStock Message Boards," *Journal of Finance* 59, 1259-1295.

Bagnoli, M., M. Beneish, and S.G. Watts, 1999, "Whisper Forecasts of Quarterly Earnings per Share," *Journal of Accounting and Economics* 28, 27-50.

Baker, W.E., 1984, "The Social Structure of a National Securities Market," *American Journal of Sociology* 89, 775-811.

Barber, B. and T. Odean, 2001, "Boys will be Boys: Gender, Overconfidence, and Common Stock Investment," *Quarterly Journal of Economics* 116, 261-292.

Blaug, M., 1992, *The Methodology of Economics*, Cambridge, Cambridge University Press.

Bushee, B., D. Matsumoto, and G. Miller, 2003, "Open versus Closed Conference Calls: The Determinants and Effects of Broadening Access to Disclosure," *Journal of Accounting and Economics* 34, 149-180.

Busse, J. and T.C. Green, 2002, "Market Efficiency in Real Time," *Journal of Financial Economic*s 65, 415-437.

Chakrabarti, S., B. Dom, R. Agrawal, and P. Raghavan, 1998, "Scalable Feature Selection, Classification and Signature Generation for Organizing Large Text Databases into Hierarchical Topic Taxonomies," *The VLDB Journal*, Springer-Verlag.

Coval, J. and T. Shumway, 2001, "Is Sound Just Noise?" *Journal of Finance* 56, 1887-1910.

Das, S. and M.Y. Chen, 2003, "Yahoo! For Amazon: Sentiment Extraction from Small Talk on the Web," Santa Clara University Working Paper.

Das, S. and J. Sisk, 2005, "Financial Communities," *Journal of Portfolio Management* 31, 112-123.

Delong, J.B., A. Shleifer, L. Summers, and R.J. Waldman, 1990, "Positive Feedback Investment Strategies and Destabilizing Rational Speculation," *Journal of Finance* 45, 379-395.

Fama, E., 1965, "The Behavior of Stock Market Prices," *Journal of Business* 38, 34-105.

Fama, E., 1972, *Foundations of Finance*, New York, NY, Basic Books.

Fleming, M. and E.M. Remolona, 1999, "Price Formation and Liquidity in the US Treasury Market: The Response to Public Information," *Journal of Finance* 54, 1901-15.

Gay, G., J.R. Kale, R.W. Kolb, and T.H. Noe, 1994, "(Micro) Fads in Asset Prices: Evidence from the Futures Market," *Journal of Futures Markets* 14, 637-59

Gervais, S. and T. Odean, 2001, "Learning to be Overconfident," *Review of Financial Studies* 14, 1-27.

Goetzman, W., M. Massa and G. Rouwenhorst, 2004, "Behavioral Factors in Mutual Fund Flows," Yale University Unpublished Manuscript.

Godes, D. and D. Mayzlin 2004, "Using Online Conversations to Study Word-of-Mouth Communciation," *Marketing Science* 23, 545-60.

Kim, O. and R.E. Verrecchia, 1991, "Trading Volume and Price Reactions to Public Announcements," *Journal of Accounting Research* 29, 302-321.

Koller, D. and M. Sahami, 1997, "Hierarchically Classifying Documents using very Few Words," International Conference on Machine Learning 14, San Mateo, CA, Morgan-Kaufman.

Lakonishok, J., A. Shleifer, and R.W. Vishny, 1994. "Contrarian Investment, Extrapolation, and Risk," Journal of Finance 49, 1541-1578.

Lee, C., A. Shleifer, and R. Thaler, 1991, "Investor Sentiment and the Closed-end Fund Puzzle," *Journal of Finance* 46, 75-109.

Leinweber, D. and A. Madhavan, 2001, "Three Hundred Years of Stock Market Manipulation," *Journal of Investing* 10, 7-16.

MacKinlay, A.C., 1997, "Event Studies in Economics and Finance," *Journal of Economic Literature* 35, 13-39.

Mitchell, M. and J.H. Mulherin, 1994, "The Impact of Public Information on the Stock Market," *Journal of Finance* 49, 923-950.

Odean, T., 1998. "Are Investors Reluctant to Realize Their Losses?" *Journal of Finance* 53, 1775-1798.

Tumarkin, R. and R. Whitelaw, 2001, "News or Noise? Internet Message Board Activity and Stock Prices," *Financial Analysts Journal* 57, 41-51.

Wysocki, P., 1999, "Cheap Talk on the Web: The Determinants of Postings on Internet Stock Message Boards," University of Michigan Unpublished Manuscript.