Social Learning and Sentiment Contagion in the Bitcoin Market

Bing Han [*] Haoyang Liu [†] Pengfei Sui [‡]

Date: January 2, 2025

Abstract

Using novel data on social interactions and individual trading records in the Bitcoin market, we document evidence of social learning leading to sentiment contagion. Investors significantly update their beliefs about Bitcoin in the same direction as the average peer sentiment, although it is not informative about future prices. Our findings indicate inefficiency in social learning, consistent with the echo chamber effect and selective interpretation of signals. Moreover, social learning affects both individuals' trading decisions and aggregate market outcomes. We construct a novel measure for the intensity of sentiment contagion resulting from social learning, which significantly predicts Bitcoin volatility, volume, and crash.

Keywords: Social Finance, Sentiment Contagion, Bubbles, Bitcoin

JEL Code: G11, G12, G41, G53

[*]University of Toronto; Bing.Han@rotman.utoronto.ca

[†]The Federal Reserve Bank of Dallas; haoyang.liu@dal.frb.org

[‡]The Chinese University of Hong Kong, Shenzhen; psui@cuhk.edu.cn

The authors thank Charles Cao, William Cong, Tony Cookson, Adlai Fisher, Byoung-Hyoun Hwang, Lawrence Jin, Weikai Li (discussant), Hongqi Liu, Yukun Liu, Song Ma, Marina Niessner, Jinfei Sheng (discussant), Tao Shu, Baolian Wang, Wei Xiong, Yao Zeng, Bohui Zhang, Dexin Zhou, and the participants at the Annual Conference on Social and Behavioral Finance, China Financial Research Conference, GSU-RFS Fintech Conference, Five-Star Asia Pacific Workshop in Finance at Singapore Management University, Central University of Finance and Economics, Cheung Kong Graduate School of Business, Fudan University, Hong Kong University of Science and Technology, Nankai University, Nanjing University, National University of Singapore, Peking University Guanghua School of Management, Peking University HSBC Business School, Renmin University of China and University of Hong Kong for their valuable comments on this paper. Part of this work is conducted while Bing Han is visiting the Chinese University of Hong Kong. An earlier version of this work was titled "Conversations in the Market: Sentiment Contagion Among Investors". All errors are our own.

1 Introduction

"*Man is by nature a social animal*"- Aristotle

Like Aristotle, researchers have long realized the importance of social activities on economic outcomes. However, the role of social interactions has been largely absent in economics and finance until recent decades (Shiller et al. (1984); Shiller and Pound (1989); Shiller (2007); Hirshleifer (2020)). In his pioneering works, Robert Shiller argues that social learning, among other factors, plays a pivotal role in influencing investors' decisions as they update their beliefs through social interactions. Shiller further proposes a potential link between social learning and asset pricing dynamics. Despite its prominence in the literature, direct empirical evidence on social learning has been limited, possibly due to data constraints.

We provide novel insights into social learning by leveraging new data on investors' social activities in the Bitcoin market. We document direct evidence for social learning and shed light on whether social learning is fully rational. We further document the aggregate impact of social learning on market outcomes. The Bitcoin market is well-suited for our study due to its distinctive social characteristics. First, Bitcoin, as a product born in the internet era, has its origins and development closely tied to social networks. Right after the inception of Bitcoin, its founder, Satoshi Nakamoto, quickly established Bitcointalk, an online social platform designed to support Bitcoin operations and facilitate information sharing and communications among Bitcoin users. Over time, Bitcointalk has grown into the most influential social platform on Bitcoin. Second, Bitcoin is a novel asset whose valuation is difficult to determine. Users often turn to social platforms to share information and views. As suggested by recent studies (Cong et al. (2021)), the valuation of Bitcoin depends on the aggregate demand which is a function of its social adoption. Thus, investors would find it useful to gather information from social media about others views' towards Bitcoin. Finally, consistent with the long-standing hypothesis on the role of social dynamics in fueling boom-bust cycles (Shiller et al. (1984), Burnside et al. (2016)), the Bitcoin market frequently exhibits high volatility and bubble-burst characteristics (Makarov and Schoar (2019)). Moreover, mistaken beliefs can form and spread via social interactions among investors (e.g., Hirshleifer (2020), Han et al. (2022)).

We utilize Bitcointalk data in our study of social learning. Bitcointalk enables investors to share opinions and engage with one another through posting and replying to each other's posts. As a result, we are able to observe conversations among investors, which is one of the most direct forms of social interactions. By leveraging Natural Language Processing (NLP) techniques, we can determine the sentiment of each message in these conversations. This setting provides an ideal environment for studying social learning. Furthermore, a subset of Bitcoin investors on this platform voluntarily disclose their Bitcoin wallet addresses to enhance the security of their forum accounts. Therefore, we can observe their wallet-level trading records on the Bitcoin blockchain. By linking the sentiment expressed in their posts to their transactions, we can draw unique conclusions about the impact of social learning on investor trading behavior and market outcomes.

We begin by examining whether and how investors learn via social interactions. The basic unit of observation in our analysis is a conversation segment that starts and ends with two posts (post[0] and post[1]) from the same investor. We use the term *social sentiment* to refer to the average sentiment of other investors' messages between post[0] and post[1]. In our setting, social learning is manifested by one's belief update (measured by own sentiment change) after being exposed to peer messages. We find that investors' sentiment move in the same direction as the social sentiment, suggesting that sentiment spreads via conversations on the social network. Specifically, investors become more bullish (bearish) after being exposed to generally positive (negative) posts from others. Such social learning or belief updating through social interactions can lead to sentiment contagion.

We address several econometric issues when interpreting our results as evidence of social learning. First, we employ a test following Altonji et al. (2005) and Oster (2019) and demonstrate that the impact of endogeneity arising from omitted variables is limited. Our finding that social sentiment significantly affects belief updating remains robust in the presence of a wide range controls. To account for common shocks that happens between the two posts that may affect both change in one's belief update and other posters' sentiment, we control for contemporaneous average news sentiment from Ravenpack News Database between consecutive posts as well as the contemporaneous Bitcoin return and volatility. We also include lagged values of these control variates to account for possible delayed reactions to common shocks.

Second, we take into consideration sample selection related to posting on social network. Our posting sample may disproportionately reflect the influence of frequent users, raising the question whether our main finding merely reflects social learning among this specific group. However, we find that the social learning effect remains strong regardless of the posting frequency. Another potential concern arising from the selective nature of participation is that the positive association between social sentiment and sentiment change may not immediately imply social learning. It is possible that individuals only continue engaging in conversations where social sentiment aligns with their initial post. If others do not agree with an individual's initial post (post[0]), the user might choose to exit the conversation. Under this alternative interpretation based on selective participation, the positive correlation between social sentiment and sentiment change would be considered ”mechanical.” To rule out this interpretation, we demonstrate that, regardless of the presence of subsequent post[1], the average sentiment in the encountered posts after post[0] significantly predicts subsequent trading decisions — a variable that is immune to selective nature of post[1].

Social learning on Bitcointalk appears not fully rational or efficient. Peer sentiments expressed in the post messages are largely uninformed. If anything, peer sentiments tend to negatively predict future Bitcoin returns. Yet, investors appear to treat the social signals as informative, and they significantly update their beliefs about Bitcoin price in the same direction as social sentiments. Further, we find that social learning is unrelated to the arrival of Bitcoin-related news in traditional media as reported by the RavenPack database. It intensifies on days with greater Bitcoin volatility or higher dispersion in the RavenPack Bitcoin news sentiment, suggesting a greater reliance on social signals during periods of high uncertainty. These results are hard to reconcile under Bayesian updating.

We uncover two patterns of inefficiency in social learning that are consistent with confirmation bias. Firstly, users selectively participate in conversations that are consistent with their priors, resulting in an echo chamber effect. For instance, users with positive prior sentiments are, on average, 6.738% more likely to engage in a thread initiated by a post with positive sentiment, compared to users with non-positive priors. To provide context, the unconditional probability of participating in a conversation that indicates a positive sentiment is 37.310%. Secondly, users selectively interpret information, responding more assertively to the social sentiment that aligns with their own prior sentiments.

To the extent that social learning is non-fully rational, we expect that sophisticated and informed investors are less affected by biased social learning. Consistent with this hypothesis, we find that investors who are less sophisticated, less socially connected, and less informed are more susceptible to social sentiment. For example, inexperienced users exhibit, on average, an additional sentiment adjustment of 3.585% after being exposed to a conversation with a social sentiment score of 1. Less informed users, compared to informed users, tend to update their sentiment by an additional 2.664% after participating in a conversation with a social sentiment score of 1. Furthermore, we consider users within a network structure and treat two users as connected if they have participated in at least one common conversation. We find that investors who have more connections are less responsive to social sentiment.

Taking a step further, we examine the impact of social learning on investors' trading decisions. We establish a clear link between the social sentiment encountered by investors and their subsequent net buying decisions using detailed trading records for a subsample of users on Bitcointalk. Our findings demonstrate that, in the presence of social interactions, social sentiment significantly predicts the direction of future trading decisions. If the encountered social sentiment increases by one unit, investors are 0.409% more likely to buy in the next day. The effect is economically significant, considering the unconditional probability of buying Bitcoin on a given day in our trading sample is 0.934%. Moreover, these transaction-based findings serve to validate our primary results about belief updating. To the extent that belief updating and trading is closely correlated, social signals that affect users' belief updating should also significantly impact users' trading decisions. Our findings confirm this hypothesis. Overall, social learning is not merely a sideshow; instead, it strongly influences individuals' transactions. At a broad level, these results establish a micro-foundation for understanding the influence of social interactions on market outcomes.

Finally, we document the aggregate impact of social learning on market dynamics. Specifically, we construct an index (SCI) to track the intensity of sentiment contagion in the market. We count the number of investors on Bitcointalk who change their sentiment in the same direction as the social sentiment in conversations and, after removing the time trend, use it as our proxy for the amount of social learning via conversations on Bitcointalk. As changes in investor sentiment are linked to individuals' future trading behaviors, we conjecture that SCI could also capture the aggregate demand for position changes in the future and, therefore, predict market dynamics such as trading volume. Consistent with our conjecture, we observe a strong predictive power of SCI for both Bitcoin trading volume and return volatility.

As sentiment spreads through the population via social interactions, more users become influenced by it. How do the accumulated belief changes via social interactions affect market outcomes? Given the documented inefficiencies in social interactions, we hypothesize that sentiment contagion may destabilize the market and contribute to the boom and bust in the Bitcoin market. We find confirmatory evidence for this hypothesis. First, we establish a link between social learning and bubbles. In our identified bubble episodes, we observe widespread optimism and a disproportionately high fraction of new investors engaging in these conversations. The contagion of optimism gets amplified in the bubble episodes and drives up the trading volume. We find that the correlation between the number of investors positively influenced by social interactions and the trading volume during bubbles can be as high as 0.669. Therefore, social learning offers a fresh perspective for comprehending the elevated trading volume during bubbles. Second, when the prevalence of optimistic views in the forum rises, the likelihood of a crash in Bitcoin returns substantially increases. This finding supports the error-prone nature of the impact of social interactions on aggregate outcomes proposed in Hirshleifer (2020).

Our paper is connected to several strands of literature. First, there is a growing interest in how investment ideas are transmitted. Shiller et al. (1984) argues that investment in assets is a social activity and Shiller and Pound (1989) considers the role of social interaction in the transmission of financial information. Han et al. (2022) offers a model of social interactions to understand the transmission pattern of investment ideas. In line with these studies, our paper provides concrete evidence for the transmission of sentiment through social learning.

Our paper contributes to the growing literature on cryptocurrencies. Makarov and Schoar (2020) and Makarov and Schoar (2019) study price formation for cryptocurrencies at the aggregate market level. In comparison, our study investigates the individual level behaviors of Bitcoin investors and then establishes meaningful connections between social learning and market dynamics. Kogan et al. (2024) find that investors seem to form adaptive expectations about cryptocurrency prices while we highlight the role of peer sentiment on investor beliefs. Liu et al. (2022a) shares a similar spirit with our paper in terms of methodology and topic, as both studies adopt a machine-learning approach to address important questions about cryptocurrencies. Liu et al. (2022a) design machine learning methods to construct technology indexes from ICO whitepapers and investigate their role in cryptocurrency valuations. In comparison, we employ machine learning tools to measure sentiment in social media and examine social learning as well as its impact on trading decisions and market outcomes. Liu and Tsyvinski (2021) evaluates the risks and returns of cryptocurrencies, and Liu et al. (2022b) documents that three common risk factors—cryptocurrency market, size, and momentum—can capture the cross-sectional expected cryptocurrency returns. In contrast, we use the Bitcoin market as an ideal laboratory to study social learning and establish significant connections between social learning and market outcomes.

Our paper closely relates to an expanding body of literature that examines investor behavior on social networks. By leveraging data from an investor social network in the foreign exchange market, Simon and Heimer (2012) document that social interaction promotes active trading. Heimer (2016) further documents an increase in the level of the disposition effect after interacting with other investors. Chen et al. (2014) highlight the usefulness of peer-based advice from the social platform SeekingAlpha for stock investing. In the context of the investor social network StockTwits, Cookson et al. (2023) document that investors tend to selectively expose themselves to information aligning with their prior views, resulting in sustained disagreement. They also link this "echo chamber" effect to high trading volume in financial markets. In comparison, we highlight sentiment contagion as a direct consequence of social learning and document how social learning predicts individuals' trading decisions and aggregate market outcomes. Additionally, we utilize social learning to account for the significant trading volume during bubbles.

Our paper is related to studies that analyze the influence of peer actions. Hong et al. (2004) and Brown et al. (2008) provide evidence consistent with the notion that individuals are more likely to participate in the stock market when their geographically proximate peers participate. Hong et al. (2005) also shows that investors tend to buy stocks their local peers have been buying in the recent past. Huang et al. (2021) document the contagion of abnormal trading activity from "infected" investors to their neighboring investors by relying on stock-financed M&A as an exogenous shock. This paper highlights social learning as one plausible driver for the peer effect. Moreover, our data allows us to directly observe and measure social interactions, which further helps establish novel connections between social learning and individual-level trading decisions. Additionally, we establish links between social learning and aggregate market outcomes.

Do peer effects "wash out" in aggregate, or do they cause significant cyclical fluctuations in asset prices? Despite the significance of this question, there have been relatively few empirical studies conducted. One notable exception is Kuchler et al. (2022), who examine the Social Connected Index based on friendship links on Facebook. Their findings reveal that investors show a greater inclination to invest in companies located in socially connected regions, and firms in regions with stronger social ties tend to have higher valuations and liquidity. In contrast, our study focuses on direct evidence of social interactions and documents sentiment contagion as a consequence of social learning among Bitcoin investors. We document several important connections between social learning and market outcomes, including future market volume, volatility, and probability of crash.

The remainder of the paper is organized as follows: Section 2 describes the data and explains the main variables used in the analysis. Section 3 documents the social learning pattern. Section 4 studies the nature of social learning. Section 5 presents the impact of sentiment contagion on individuals' trading decisions. Section 6 reports the impact of social learning on market outcomes. Section 7 concludes.

2 Data

2.1 Bitcointalk

Our study focuses on the Bitcoin market as it provides valuable opportunities to trace the full history of users' social interactive activities since its inception in 2009. To gather data on social interactions, we primarily rely on Bitcointalk, an online forum that is considered to be the oldest and most influential community in the Bitcoin space, with over 54 million messages posted for over 1.2 million topics and more than 3.5 million registered users as of March 2023. The founder of Bitcointalk, Satoshi Nakamoto, is the presumed pseudonymous author (or authors) of the original Bitcoin white paper, which describes Bitcoin's reference implementation. Bitcointalk holds significant influence on Bitcoin trading and the real world, serving as a hub for traders to discuss market trends, share trading strategies, and provide insights into the direction of the market. Its popularity has been widely covered by influential media outlets such as the Wall Street Journal, Forbes, and Bloomberg. Considering its impact on Bitcoin investors, we use social interactions on this forum as a proxy for the general trend of social activities in the Bitcoin market.

In this paper, our research is centered around posts on Bitcointalk, which are timestamped messages containing sentences written by users. Generally, sentences within each post are concise in nature. These posts are organized within threads, which are collections of sequential posts that revolve around a common theme, allowing users to interact with each other. As an example in our sample, Figure 1 presents a thread titled "What does the future of bitcoin look like" joined by user DavidLuziz on May 31, 2018, with subsequent participations from other users sharing their opinions.

On Bitcointalk, threads are displayed under different "child boards", which are subforums that gather discussions related to specific topics. Not all subforums are about topics related to the prospect of Bitcoin prices. For instance, the "Bitcoin Technical Support" child board is dedicated to addressing technical questions concerning Bitcoin Core, nodes, the Bitcoin network, transactions, and addresses. For our study, we specifically focus on the "Speculation" child board, where users predominantly discuss their views on Bitcoin's prospects. Figure 2 presents the visual representation of the keywords mentioned in posts within this child board. The most frequent words include "Bitcoin", "price", "think", and "know". These keywords indicate that the topics are quite concentrated, and investors express their attitudes about Bitcoin prices in this subforum. The "Speculation" child board serves as an ideal field laboratory for our research, as we can examine how social interaction propagates users' sentiment, and how sentiment contagion is linked to individual trading behavior and aggregate market outcomes.

There are multiple layers of user heterogeneity on this forum. For instance, we observe variations in the level of sophistication among users. Bitcointalk utilizes a merit system wherein users are awarded points for their contributions to the Bitcoin community. Based on their merits and activities, users are assigned different ranks on the Bitcointalk website, ranging from high ranks such as "Legendary" to lower ranks like "Newbie". Typically, users with higher ranks demonstrate a greater understanding of Bitcoin. Moreover, a significant number of users voluntarily disclose their demographic information, including age, gender, and home country.

2.2 User Sentiment

2.2.1 Textual Analysis and User Sentiment

In this paper, "user sentiment" refers to users' beliefs about future Bitcoin prices. We develop a methodology to extract user sentiment from posts on the Bitcointalk forum. Since our raw data consists of over 1 million posts, far beyond what we can manually interpret, we employ a textual analysis algorithm. Our algorithm is a two-step procedure based on a keyword dictionary and a natural language processing (NLP) algorithm developed by the Stanford University NLP group (Manning et al. (2014)).

In the first step, we randomly select 10,000 sentences from our dataset and manually label them into four categories: "positive", "neutral", "negative", and "irrelevant". This approach mainly follows the methods used in Baker et al. (2016) and Tetlock (2007). We then use these labeled sentences as our training set and construct a keyword dictionary for each category. For example, keywords under the "positive" category include "buy", "increase", and "rise", while the "negative" category contains phrases such as "sell", "decrease", and "plunge", and so on. Additionally, we process sentences containing negative particles like "not" and "couldn't". If a sentence with negative particles contains words that are classified as "positive", we reverse our interpretation and label the sentence as "negative" instead of "positive". The dictionary for the "neutral" category includes terms such as "hold", "wait", "unpredictable", and others. If a sentence does not contain any of the keywords in the above three categories, we label it as "irrelevant".

In the second step, we apply the Stanford NLP algorithm to detect sentences that describe the future. One challenge we face in our labeling exercise is distinguishing between descriptive statements about past performance and beliefs about the future in certain posts. For instance, a sentence like "Bitcoin market really increased a lot in the past months" can be ambiguous in terms of an user's opinion about future Bitcoin prices. Relying on the Stanford NLP algorithm, we can identify sentences with forward-looking statements, and label the rest as "irrelevant". For example, if a sentence contains the keyword "increased", the Stanford NLP would detect the past tense, and our algorithm would label it as "irrelevant". On the other hand, the Stanford NLP algorithm would identify future tense in sentences with phrases such as "will increase". Furthermore, our definition of forward-looking statements aligns with the format recommended by the SEC for 10-K reports, including sentences with terms like "expects", "anticipates", and so on.[1]

After the aforementioned two steps, we extract user sentiment from the relevant sentences. We assign a value of 1 to sentences with a "positive" sentiment, -1 for "negative", and 0 for all others. For example, a sentence like "Bitcoin price will roar" is assigned a value of 1, while "Bitcoin price is not going to fall" is also assigned a value of 1. Conversely, a sentence like "Bitcoin is doomed to fall" is assigned a value of -1, and a sentence like "Bitcoin price is really unpredictable" takes a value of 0. To calculate the overall sentiment of a post, we take the average of the sentiment values of all its sentences. As a result, the sentiment measurement for a post falls within a continuous range from -1 to 1. The out-of-sample accuracy of our algorithm is approximately 85%.

[1] See $https://www.sec.gov/Archives/edgar/data/1082027/000139390519000101/neik_10k.htm$ for details.

2.3 Transaction Data

We are able to link the sentiment of a subsample of users to their trading behaviors. This group of users voluntarily published their Bitcoin wallet addresses in a thread officially initiated by the forum organizer, as a protective measure for their Bitcointalk accounts against hacking.[2] Since Bitcoin transactions are publicly available on the Bitcoin Blockchain, we can trace their detailed transactions information such as trade size and trade timestamp using these published wallet addresses.

2.4 Other Data Sources

To proxy investor attention for the Bitcoin market, we utilize Google search volume data following Da et al. (2011). Specifically, we define Google Search (Bitcoin) as the difference between the Google Search Volume Index and its past one-month mean, divided by the lagged one-month mean.

To systematically capture the impact of news events on the Bitcoin market, we obtain news related to Bitcoin from Ravenpack News Analytics, a reliable source that tracks news reports about Bitcoin and provides sentiment scores for them dating back to 2011. We refer to the general tone of news coverage on Bitcoin in the media as the news sentiment, and include it as a control variable in our analysis.

We gather Bitcoin market data, including returns and trading volume on an hourly basis from CoinAPI. We also calculate daily return volatility. This data is also utilized in Griffin and Shams (2020). The transactions before May 2012 were relatively less frequent. Therefore, our focus is on transaction data starting from May 1st, 2012, and the sample period ends on July 30th, 2022.

2.5 Summary Statistics

To be consistent with the Bitcoin market data, our main dataset from Bitcointalk covers the period between May 1st, 2012, and July 30th, 2022. It comprises 666,006 posts in the ”Speculation” subforum,

[2] This is an official activity organized by the management team of the Bitcointalk forum. Users of Bitcointalk are recommended to disclose their public Bitcoin wallet address within this thread. When a user posts their public Bitcoin wallet address on Bitcointalk, they are creating a public record of their ownership of that wallet address. If their Bitcointalk account is later hacked, they can use their private key to create a digital signature that proves their ownership of the Bitcoin address they previously shared. This digital signature can then be used to authenticate the account holder’s identity and facilitate the recovery of their Bitcointalk account.

contributed by 44,356 users. Among these users, we successfully linked 2,550 individuals to their transaction records through their disclosed Bitcoin wallet addresses. In Panels A, B, and C, we present summary statistics for user sentiment at the user, daily, and thread levels, respectively. Panel D reports summary statistics for market information of Bitcoin, RavenPack news sentiment and Google Search Volume.

In Panel A of Table 1, we present the summary statistics for posting activities at the user level. For a representative user in our sample, the average sentiment score of the posts they published is 0.324, suggesting a general optimism towards Bitcoin among users. Such optimism is consistent with the existing literature on motivated beliefs (Bénabou and Tirole (2016)) and utility-based biases in beliefs (Brunnermeier and Parker (2005)). Furthermore, we observe a significant within-user standard deviation of sentiment at 0.608, indicating substantial variation in sentiment over time at the individual level. We require at least two posts for a given user to compute the standard deviation of post sentiment. The distribution of the number of posts is skewed, as evident from the difference between the median and mean number of posts per user.

Panel B of Table 1 describes post activities on a given day. On a representative trading day, about 120 users publish around 178 posts. The average post sentiment is 0.279, which is optimistic and consistent with the statistics at the user level.

Panel C reveals that posts within threads often exhibit dispersed sentiment, with an average within-thread standard deviation of 0.646. Furthermore, active user engagement is notable, with 25 users participating in a representative thread containinga total of 35 published posts. Overall, users seem to pay close attention to the posts, resulting in active exchanges within threads, with a median gap between consecutive posts within a thread of only 48 minutes.

Panel D provides summary statistics for variables at the aggregate market level. In the first row, the mean return (annualized) is 119.0% and the standard deviation is 16.166. When combined, the Sharpe ratio is approximately 0.074. The second row presents the within-day volatility of Bitcoin, calculated using hourly return data. On average, the volatility is approximately 3.7%. We present volume-related variables for Bitcoin, spanning from the third to the fifth row. On a typical trading day, there are 16,472 transactions involving 9,359 bitcoins being traded. The average daily trading volume totals 56.499 million dollars. Moving to the sixth row, we present summary statistics for the sentiment of RavenPack News. In its original form, the sentiment score for RavenPack News Database ranges from 0 to 100, but we have normalized it to the range of $[-1, 1]$. On average, the transformed sentiment of news within a single day is 0.041 with a standard deviation of 0.330. The daily correlation between RavenPack News sentiment and post sentiment is 0.262, indicating that while there is some overlap, these two data sources capture distinct dimensions of sentiment.

3 Social Learning and Sentiment Contagion

In this section, we investigate whether social learning propagates the spread of peer sentiment, a pattern we refer to as *sentiment contagion*. Specifically, we examine how a user's sentiment may be influenced by the sentiment of others, as expressed through conversations on the forum.

3.1 Data Structure

Figure 1 illustrates an example of conversation from the Bitcointalk forum. User DavidLuziz published one post at 05:37:35 AM on May 31st, 2018 (post[0] at the top). Subsequently, several other users joined the conversation and shared their views about Bitcoin. At 07:50:10 AM on the same day, DavidLuziz published another post in the same thread (post[1] at the bottom), becoming more optimistic about Bitcoin after interacting with other users.

We study consecutive posts from the same user, similar to the ones by DavidLuziz in Figure 1. We denote the first post published at timestamp t_0 as post[0], and the second post made at timestamp t_1 as post[1], as illustrated in Figure 3. We refer to the sentiment in post[0] and post[1] as the *prior sentiment* and the *posterior sentiment*, respectively.

Between a consecutive pair of posts by a user, other users may publish posts, forming what we term as a *conversation*. To measure the views of peers in social interactions, we define *social sentiment* as the average sentiment of other users' posts within the conversation. One user's consecutive posts could be in two different threads.[3] In such cases, conversations refer to posts published between t_0 and t_1 in both threads to which post[0] and post[1] belong. Our results remain robust to restricting to a smaller sample with post[0]

[3] After posting in one thread, the user will receive notifications from the forum if others publish follow-up posts in the same thread. Therefore, users will pay attention to subsequent posts in the thread to which post[0] belongs.

and post[1] in the same thread (see Appendix Table A4). In total, there are 296,291 pairs of posts by the same user (and correspondingly 296,291 conversations) from 16,535 users in our main test sample.

Figure 3 illustrates the timeline of our analysis. There is a trade-off when determining the time window allowed between the consecutive pairs of posts. A shorter time window would reduce the confounding influence of unobservable factors for belief changes, but it would result in a smaller sample of conversations, thereby reducing the statistical power of our tests. We strike a balance by choosing a window that is sufficiently short in order to cleanly capture belief changes caused by social interactions while avoiding excessive data loss. Figure 4 presents the cumulative distribution function of the time gap between consecutive posts by the same user in our sample. More than 50% of the pairs fall within a 24-hour window. Therefore, we select a 24-hour time window in our main analysis.[4]

3.2 Empirical Strategy

To test sentiment contagion via social interactions on Bitcointalk, we examine whether and how a user updates her belief about Bitcoin after being exposed to others' views in the conversation. We achieve this by linking the user's sentiment change, defined as the revision from her prior to posterior post sentiment, to the social sentiment. We control for confounding factors that might affect sentiment change, such as Bitcoin return and volatility, common news and the overall sentiment of contemporaneous posts on the forum. Our main regression model is:

$$\begin{aligned}\text{Senti Change}_{i,j,t_0\to t_1} &= \beta_1 \text{Social Sentiment}_{i,j,t_0\to t_1} \\ &+ \gamma' \text{Control}_{i,t_0\to t_1} + \text{Fixed Effects} + u_{i,t_1}, \end{aligned} \tag{1}$$

where Senti Change$_{i,j,t_0\to t_1}$ is the revision in user i's sentiment from post[0] at t_0 to post[1] at t_1 during conversation j.[5] Social Sentiment$_{i,j,t_0\to t_1}$ is the social sentiment that user i has encountered in conversation j between time t_0 and t_1. We include a set of control variables, denoted as Control$_{i,t_0\to t_1}$, along with various fixed effects to address several concerns related to issues such as endogeneity. We report standard errors that are clustered at the user and day levels.

[4] Appendix Table A5 indicates that the effect of social sentiment on belief updating is even stronger for 48 hours or 72 hours time window than 24 hours window.

[5] Appendix Table A1 verifies the robustness of our findings when we use the posterior sentiment as the dependent variable.

We choose to use sentiment change as the primary dependent variable because we consider it a more direct measure of the social learning effect. This variable also aligns more coherently with our subsequent analysis of the impact of social learning on market outcomes. Another option for the dependent variable would be the ex-post sentiment. Appendix Table A1 verifies the robustness of our findings under this alternative setting. In fact, the magnitudes become larger when we use ex-post sentiment as the outcome variable, indicating that we provide a conservative estimation of the effect of social learning under the current setting using sentiment change. Moreover, our regression models incorporate fixed effects to account for the signs of the prior sentiment (hereafter prior fixed effect). This choice is motivated by the observation that users with a positive prior sentiment naturally have less room for increased optimism compared to those with a negative prior sentiment.

3.3 Evidence for Social Learning

The baseline regression in Table 2 column (1) shows a significantly positive relation between sentiment change and social sentiment with a $t - stat$ of 6.81 for the regression coefficient of social sentiment. A one standard deviation increase in social sentiment is associated with an increase in changes in sentiment of 1.197% (calculated as $42.177\% \times 2.838\%$). The effect is significant economically, given the average sentiment level of 0.324.

Columns (2) to (5) show that the effect of social sentiment remains significant and robust after we include various controls to address potential concern about omitted variables. Column (2) adds the average news sentiment from Ravenpack News Database between consecutive posts to control for contemporaneous common news arrivals (Feng and Seasholes (2004)). In column (3), we include the Bitcoin return and volatility between t_0 and t_1. In column (4), we further control for the contemporaneouus forum sentiment, computed as the average sentiment of posts published between t_0 and t_1 in all conversations on Bitcointalk other than those involving post[0] or post[1]. Furthermore, to address concerns that investors may respond to news with a delay, we include in column (5) lagged Ravenpack news shocks up to 48 hours before post[0] as additional controls.[6] In Appendix Table A6, we further include Bitcoin return and volatility within past 24- and 48-hour windows, as well as the average sentiment of posts before post[0], to account for investors'

[6] In untabulated results, we also control for lagged news arrivals up to the past 7 days, and our findings remain consistent.

delayed response to market dynamics and social interactions. We find that the regression coefficient of social sentiment remains significantly positive in all cases. We also verify the robustness of our results when user-week fixed effects are included (so that we solely rely on variation between different post pairs by the same user in the same week), suggesting that our results are not driven by time-varying user characteristics.

3.4 Placebo Test

To further establish the significance of social learning, we conduct a placebo test. Specifically, we investigate whether sentiment changes are influenced by the average sentiment of randomly selected conversations that take place between the timestamps of post[0] and post[1] but in which the user does not participate. The results are presented in Table 3. In column (1), we randomly select one conversation and find that the average sentiment from this conversation does not significantly predict sentiment change. The coefficient magnitude is only 5% of that observed for social sentiment in Table 2. To demonstrate robustness, we also examine the effect of selecting multiple random conversations. For example, in column (4), we randomly select ten conversations that occur between post[0] and post[1], in which the user does not participate. We find that the non-significant pattern remains robust.

3.5 Endogeneity

Our tests face potential endogeneity issues. One concern is reverse causality, which appears unlikely in our setting, as we examine whether one's belief is updated significantly after being exposed to others' views in a conversation. Another source of potential endogeneity is an "omitted variable bias" (Jiang (2017)). For example, the decision to participate in a conversation could be influenced by private motivations that are unobservable to econometricians. It is also possible that the belief updating of a user from post[0] to post[1] and the sentiment of other posts in between are both affected by some unobserved factor such as common news arrival. However, one significant advantage of our analysis is that we can observe users' social activities at the timestamp level. This enables us to construct granular concurrent factors of high-frequency, such as news arrivals and market return. For example, we include the average news sentiment from Ravenpack News Database between the two consecutive posts as a regressor to control for the effect of contemporaneous news arrivals (Feng and Seasholes (2004)). We further control for the Bitcoin returns between the two consecutive posts in order to soak up the possible effect of common shocks that happens between post[0] and [1]. Another source of omitted variables arises from users' lagged responses to events that occur before the conversation starts. Since we observe the exact time stamp when a conversation starts, we are able to address this issue by including lagged news arrival, lagged market return, and lagged social posting activities etc.

Despite our inclusion of a set of nuanced controls to mitigate the omitted variable issues as much as possible, concerns may still arise regarding the controls' imperfections in fully addressing endogeneity issues. To delve deeper and evaluate the effect of omitted variable bias, we conduct a series of tests following Oster (2019). The intuition behind the test is that it is important to interpret movements in coefficients and R^2 jointly. If an existing control significantly improves R^2 without moving the coefficient much towards 0, then the estimated effect is probably not due to an omitted quality variable. We can extrapolate and infer that adding more controls will not change the results to be statistically insignificant.

The formal test incorporates the change in R^2 resulting from the addition of controls. To conduct this test, we first estimate the linear model presented in equation (1) without any control variables, allowing us to obtain the coefficient of social sentiment β_u and the coefficient of determination R_u^2. Then we include contemporaneous and lagged control variables (see the last column of Table 2) to derive the coefficient β_c and the coefficient of determination R_c^2.

Oster (2019) shows that under some technical condition, the null hypothesis of no relation can be tested by checking whether zero belongs to an identified set of coefficients defined as the interval between β_{adj} and β_c, where the bias-adjusted coefficient β_{adj} is given by

$$\beta_{adj} \approx \beta_c - \delta \frac{(\beta_u - \beta_c)(R_{max}^2 - R_c^2)}{R_c^2 - R_u^2}.$$

The identified set of coefficients relies on two crucial parameters: δ and R_{max}^2. The parameter δ quantifies the level of selection on unobservable factors relative to observable controls. A higher value of δ indicates a greater severity of the omitted variable problem. On the other hand, the parameter R_{max}^2 represents the hypothetical overall R^2 of the model when both observable and unobservable variables are accounted for. It indicates the extent to which the variation in the outcome variable can be explained by controlling for all relevant factors. An upper bound on R_{max}^2 is 1. Oster (2019) argues that a reasonable case to consider is $\delta = 1$ and $R_{max}^2 = 1$. In our analysis, we experiment with different parameter values.

Table 4 presents the results of the analysis. In the upper left panel, we follow Mian and Sufi (2014) and use $R^2_{max} = \min(2.2R^2_c, 1)$ and $\delta = 1$. The adjusted β value is 0.083, allowing us to reject the null hypothesis that $\beta = 0$. Moving to the right panel, we adopt a more aggressive stance with $\delta = 2$, corresponding to the unrestricted estimator in Oster (2019). Despite this, the identified set is $[0.031, 0.177]$ which remains significantly distant from zero. In the bottom panel, we set R^2_{max} to be 1, and regardless of whether δ equals 1 or 2, we can still reject the null hypothesis. Overall, results in Table 4 show that across a wide range of scenarios, omitted variables would not qualitatively affect the significance of the positive relationship between social sentiment and sentiment change.

3.6 Selection Issues

We acknowledge potential selection issues in posting messages on Bitcointalk. Since participating in a conversation is voluntary, one concern is that individuals may be more inclined to sustain their involvement in conversations that align with their existing sentiment. When the follow-up messages mostly differ in sentiment from an individual's initial post (post[0]), the user might choose to exit the conversation (i.e, there is no post[1]).[7] Such scenarios are not included in our tests (because we need post[1] in order to measure updating in user's belief). If there are no social learning in such scenarios (e.g., people who refrain from posting messages to debate with different views are stubborn or confident about their own prior so they do not change their views in the presence of social signals), then our test would produce an over-estimation of the strength of social learning. However, the sample selection issue could not explain our evidence of social learning for the sample where we observe both post[0] and post[1]. Thus, the sample selection issue above could affect our results quantitatively but not qualitatively.

We provide additional evidence in support of social learning in a subsample that is not subject to the selective posting issue above. Specifically, we leverage the users' trading data to measure change in belief. For users who voluntarily reveal their wallet address[8], their trading records are observable. For posts by these users, we measure social sentiment as the average sentiment of posts within a subsequent 24-hour window after post[0], but we do not require the presence of a follow-up post[1]. Instead of linking social

[7]A plausible explanation for such behavior is "emotion regulation". For an excellent discussion of this concept, please refer to Hwang (2023).

[8]For detailed discussions of this subgroup of users, please see sections 2.4 and 5.

sentiment to subsequent sentiment change, we examine whether social sentiment positively predicts future trading decisions. This is confirmed in the data (see Section 5 and Table 12 for details), supporting the idea that social signals significantly impact users' belief updating and trading.

We also examine whether our main finding merely reflects social learning among frequent posters because our sample observations come disproportionately from users who choose to post messages frequently. We categorize users based on their posting frequencies into two subsamples and rerun our main regressions separately. Specifically, we calculate the average time interval between users' consecutive pairs of posts. We define users with an average interval above the population average as infrequent posters, and the remaining users as frequent posters. Our findings are presented in Table 5. Columns (1) and (2) indicate a strong and significant social learning effect for both the frequent and infrequent users. Furthermore, Column (3) shows that there is no significant difference in the magnitude of social learning effect between frequent and infrequent users.

Finally, it is worth noting that social network in the real world is unlikely to be fully exogenous and non-selective. We take as given the social interactions on Bitcointalk and ask: are users' beliefs affected by others' opinions, despite potential selection biases of the posted messages? Our evidence is based on people who posted messages on Bitcointalk. Since users of Bitcointalk are only a fraction of the overall population, therefore caution is needed to extrapolate our findings to the general population. However, as documented later in Section 6, variables we extract from social interactions on Bitcointalk have a significant impact on various bitcoin market outcomes, suggesting that the posters we study are consequential as a group and Bitcointalk is not a sideshow.

4 Non-Fully Rational Learning: Suggestive Evidence

In this section, we evaluate the nature of social learning. The traditional view is that individuals update their beliefs in a rational manner by adhering to Bayes' theorem. However, recent evidence suggests non-fully rational learning. Individuals may assign disproportionate weight to information with certain features or selectively acquire and interpret information. They may also react to peer sentiment and update their beliefs based on messages from social media, even when social signals are not informative or not new. While our objective is not to exclude Bayesian learning, our analysis highlights the existence of non-fully rational components within the social learning process, which align more closely with a behavioral narrative.

4.1 Heterogeneity by User Features

To shed light on the nature of social learning, we start by investigating which types of users are more susceptible to it. In asset pricing models, investors are typically classified into two distinct groups based on their level of rationality: naïve investors, who are significantly influenced by psychological biases, and sophisticated investors, who exhibit more rational behavior and act as the counteracting force to market sentiment (De Long et al. (1990); Lee et al. (1991); Barber and Odean (2013)). Therefore, our initial focus is on examining how social learning differs between naïve users and sophisticated users. On Bitcointalk, users are classified into legendary and non-legendary categories based on their influence and contribution to the Bitcointalk community. We adopt this classification method and designate non-legendary users as naïve users, while treating legendary users as sophisticated users. Column (1) of Table 6 presents the heterogeneous effect of social learning among these two groups. The interaction term between the naïve user indicator and social sentiment is positive and significant, suggesting that naïve users are more responsive to social sentiment compared to sophisticated users. On average, when faced with a social sentiment of 1, a naïve user updates sentiment by 3.585% more than a sophisticated user.

Second, we compare users who are more central in the social network with less central users. Central users are those who are more connected to other users. We treat two users as connected if they have participated in at least one conversation together. Central users engage widely in conversations and are likely to be more experienced in social activities. Following an established algorithm in the network literature (Hagberg et al. (2008)), we calculate the centrality score for each user on a daily basis. Column (2) of Table 6 shows that in general, central users exhibit significantly lower response to social sentiment. A one-unit increase in the centrality score reduces the magnitude of sentiment contagion by 0.047%. The social learning effect is moderated for central users but it does not get eliminated. For instance, even for a user at the 75th percentile with a centrality score of 16.3, the effect of social learning remains at 3.320% (calculated as 4.086% - 0.047% * 16.3).

Third, we investigate whether users with distinct levels of informedness respond differently to social sentiment. On a rolling basis, we define less informed users as those whose sentiment exhibits a low correlation with future Bitcoin returns in 7 days (correlation $<$ 25th percentile). Column (3) of Table 6 presents evidence consistent with this hypothesis. The interaction term indicates that, compared to more informed users, less informed users update belief by 2.664% more after participating in a conversation with a social sentiment of 1.

Overall, there is substantial heterogeneity regarding social learning. Less sophisticated, less experienced, and less informed investors—those who are likely less rational —are more responsive to social sentiment. These findings provide suggestive evidence for the existence of non-fully rational components in social learning.

4.2 Time-Varying Social Learning

To further understand the nature of social learning, Table 7 examines how social learning varies over time. First, we investigate whether social learning is driven by information flow. We identify a day as informative if the total number of news arrivals in the RavenPack News database exceeds its sample median.[9] We expect to see a stronger effect of social learning on informative days if posts on Bitcointalk mainly serve to propagate news. Column (1) of Table 7 indicates a statistically insignificant coefficient for the interaction term of social sentiment and the dummy for informative days, suggesting that social learning is not stronger when there is more new information. Therefore, social learning does not seem to be driven by the flow of information.

Second, we investigate whether social learning is influenced by the level of uncertainty in the Bitcoin market. Tversky and Kahneman (1974) established that availability heuristics influence judgments made under uncertainty. In our context, during periods of heightened uncertainty, availability heuristics may prompt individuals to be more inclined to conform to the social sentiment they encounter. Column (2) of Table 7 shows that social learning is significantly stronger when the standard deviation of news sentiment on Bitcoin from the RavenPack News database is higher than the sample median. In column (3), we proxy for uncertainty using the standard deviation of hourly Bitcoin returns in a rolling 24-hour window and document a similar pattern. Overall, we provide consistent evidence for elevated social learning in uncertain periods.

[9] In our untabulated analysis, we verify that our findings remain the same when we classify a day as informative if the total number of "novel" news labeled by RavenPack exceeds its sample median.

4.3 Informativeness of Social Sentiment

4.3.1 Social Sentiment and Future Returns

To further understand the rationality of social learning, we estimate the relationship between social sentiment and future returns by running the following regression:

$$\begin{aligned} \text{Cumulative Return}_{i,j,t_1+1\rightarrow t_1+k} &= \beta_1 \text{Social Sentiment}_{i,j,t_0\rightarrow t_1} \\ &+ \gamma \text{Control}_{t_0\rightarrow t_1} + \text{Fixed Effects} u_{i,t_1}, \end{aligned} \tag{2}$$

where the dependent variable, denoted as Cumulative Return$_{i,j,t_1+1\rightarrow t_1+k}$ represents the ex-post cumulative return in a forward-looking window from hour 1 to hour k after user i has participated in a conversation j that ends at timestamp t_1. We study five different time windows, 6, 24, 48, 72 and seven days after the conversation. The main explanatory variable, Social Sentiment$_{i,j,t_0\rightarrow t_1}$, represents the social sentiment encountered by user i in conversation j between timestamps t_0 and t_1. To address potential confounding factors, we further include a set of time-varying controls. We also include user and date fixed effects as well as the prior fixed effect. Standard errors are clustered by user and day, as recommended by Hodrick (1992).

This specification follows a well-established approach in the retail trading literature (Barber and Odean (2000)). The retail trading literature assesses the quality of decision-making by retail traders based on the timing of buys and sells relative to subsequent returns. Similarly, we evaluate the quality of information linked to social sentiment by analyzing its timing in relation to subsequent Bitcoin returns. In this context, a positive coefficient for β_1 would indicate that social sentiment is informative and enhances the quality of users' signals. Conversely, a negative coefficient indicates low informativeness, suggesting that social sentiment undermines the quality of users' signals. Table 8 shows that social sentiment is largely misinformed. Social sentiment negatively predicts future returns in most horizons. If users follow a trading strategy of buying bitcoin immediately after observing a social sentiment of 1 and hold for the next 24 hours, on average they would incur a loss of -0.094%. In comparison, Table 1 shows that a simple buy-and-hold strategy yields a daily return of 0.326% (calculated as 1.190/365).

4.3.2 Social Learning Through Informative Posts

While, on average, a higher social sentiment predicts lower returns, there may still be informative posts that positively forecast returns. Users may respond more to these informative posts. To explore this hypothesis, we analyze three user features discussed in Section 4.1 and Table 6: the level of sophistication represented by the legendary status, centrality, and post informativeness. For each highlighted feature, we separately measure the social sentiment by featured users and by non-featured users. Subsequently, we investigate whether each group's sentiment predicts future Bitcoin returns.

Panel A of Table 9 presents our findings. In columns (1) and (2), we compare the informativeness of social sentiment from legendary and naive users. Although social sentiment from naive users is slightly less informative than that from legendary users, both negatively predict future cumulative returns in the next 24 hours. Columns (3) and (4) demonstrate that the social sentiment of both central and non-central users negatively predicts returns. Somewhat surprisingly, social sentiment from central users more strongly predicts negative returns than social sentiment from non-central users. Columns (5) and (6) show that social sentiment by both informed and less informed users negatively predicts returns, although social sentiment by informed users seems relatively more precise.

In Panel B of Table 9, we investigate how users respond to social sentiment from users with different features. In columns (1) and (2), we find that users are twice as sensitive to social sentiment from naive users compared to social sentiment from legendary users. This finding is difficult to reconcile with a rational narrative, as social sentiment from both naive and legendary users negatively predicts future returns. As shown in columns (3) and (4), users strongly respond to the social sentiment from both central and non-central users. However, the social sentiment from both types of users negatively predicts future returns, as shown in columns (3) and (4) of Panel A. Columns (5) and (6) indicate that users respond positively to social sentiment from non-informed users and informed users, but this social sentiment also negatively predicts future returns. In untabulated tables, we include the user-week interactive fixed effect and find that our results remain similar. Overall, the findings from Panel A and B of Table 9 provide additional support for the behavioral narrative of social learning: different components of social sentiment are generally uninformative, but users actively learn from them and revise their sentiment accordingly.

4.4 Confirmation Bias

Why would users update their beliefs after being exposed to non-informative social signals? They might believe in wisdom of crowd, not realizing the lack of the information content of social sentiment. The psychology literature has extensively documented humans' cognitive limitations, which affect our perception, attention, memory, decision-making, and problem-solving abilities. One notable heuristic is confirmation bias (Nickerson (1998)), the tendency to search for, interpret, favor, and recall information in a way that confirms or supports one's prior beliefs or values. In this section, we study confirmation bias in social learning. We provide two pieces of evidence consistent with confirmation bias. First, users selectively participate in confirmatory conversations, a phenomenon also known as *echo chambers*. Second, they also selectively interpret the signals they encounter in conversations.[10]

4.4.1 Echo Chambers: Selective Participation of Conversations

Confirmation bias may cause users to selectively acquire signals. We examine users' decisions to participate in conversations that are confirmatory with their priors by estimating the following probit model:[11]

$$
\begin{aligned}
& Pr[\text{Participate Positive}_{i,t+k} = 1]_t \\
= \;& \Phi[\beta_0 + \beta_1 \text{Prior Sentiment}_{i,t} + \gamma_m \text{Control}_{i,t,m} + u_{t+k}], \qquad (3)
\end{aligned}
$$

where Φ is the $c.d.f.$ of the standard normal distribution. The dependent variable Participate Positive is an indicator, taking a value of 1 if the subsequent conversation after post[0] has a positive sentiment and 0 otherwise. We employ two measures for whether the subsequent conversation has a positive sentiment. The first measure is based on the first post of the thread to which post[1] belongs since it garners the most attention when users browse through threads. The second measure takes a value of 1 if the social sentiment (i.e., the average sentiment of posts between post[0] and post[1]) is positive and 0 otherwise. The explanatory variable, Prior Sentiment$_{i,t}$, is an indicator for whether user i's sentiment is positive in her post[0] published at the timestamp of t. We also include a battery of control variables Control$_{i,t,m}$ to account for market and forum events that occur between the two consecutive timestamps t and $t+k$. The control variables are

[10] By designing survey experiments, Faia et al. (2022) demonstrate how individuals selectively acquire and interpret information during the Covid-19 pandemic.

[11] The consecutive pairs of posts enable us to observe both users' prior beliefs and their subsequent participation decisions.

identical to those in Table 2.

Echo chambers predict that users with positive priors would more likely choose to participate in conversations signaling positive sentiment. We would therefore anticipate a positive coefficient for the Prior Sentiment$_{i,t}$ from estimating equation (3). The marginal effects reported in Table 10 corroborate the existence of echo chambers. For example, in column (1), the interpretation of the reported marginal effects is as follows: users with positive priors, on average, are 6.738% more likely to subsequently participate in a thread that starts with a post with positive sentiment, compared to users with non-positive priors. The unconditional probability of participating in a conversation that starts with a positive sentiment is 37.310%. Thus having a positive prior increases the probability of participating in a thread with a positive first post by 18.060% (6.738/37.310). The marginal effects of positive priors are robust to additional controls included in column (2). In columns (3) and (4), we estimate the regression with an alternative dependent variable that equals 1 when the social sentiment is positive and 0 otherwise. The marginal effects of positive priors remain significant at the 1% level.

4.4.2 Selective Interpretation of Information

Confirmation bias can lead users to interpret information in a way that confirms their priors. To examine this phenomenon, we investigate how users selectively interpret signals they encounter in conversations. Specifically, we create two new variables: social sentiment (-) and social sentiment(+). Social sentiment (-) is defined as the minimum value of social sentiment and zero. Social sentiment (+) is defined as the maximum value of social sentiment and zero. These two variables capture the linear relationship between sentiment contagion and social sentiment in the negative and positive regions, respectively. We then test for selective interpretation using two separate subsamples: one for users with positive priors and one for those with non-positive priors. Confirmation bias predicts that users with positive priors would respond more strongly to social sentiment (+), while users with negative priors would respond more aggressively to social sentiment (-).

Table 11 presents our findings. In column (1), we observe that users with positive priors adjust their sentiment towards the social sentiment, which supports the main message conveyed in Table 2. Column (2) shows that users with positive priors respond to both positive and negative social sentiment, but they are much more responsive to positive sentiment. Specifically, a one-unit increase in the positive social sentiment is associated with a 5.012% probability (statistically significant at 1% level) of a sentiment upward change. In contrast, a one-unit increase in social sentiment when it is negative is associated with only 2.687% probability (which is not significant at 10% level) of own sentiment change. In column (3), we find that users with negative priors also adjust their sentiment towards the social sentiment. Finally, in column (4), we observe that users with negative priors respond to both positive and negative social sentiment, but they are more responsive to negative social sentiment. In untabulated tables, we include the user-week interactive fixed effect and find that our results remain similar. Taken together, the results in Table 11 indicate a selective interpretation of information: investors respond more aggressively to the social sentiment that is consistent with their prior sentiment.

5 Social Learning and Individual Trading

In this section, we provide micro-level evidence on the link between social learning and individuals' trading decisions for a subsample of users on Bitcointalk. Out of the 44,356 registered users in our sample, 2,550 users voluntarily published their Bitcoin wallet addresses, following the recommendation of the Bitcointalk website (explained in Section 2.3). Since Bitcoin transactions are public on the blockchain, we have access to the complete transaction history of these 2,550 users.

Importantly, from the perspective of social learning, users within this subsample are not special. Regardless of whether users choose to reveal their wallet address, the main effect we document in Table 2 remains robust. We present the detailed results in Appendix Table A7. The intensity of social learning exhibits no significant difference between the two subsamples, as shown in column (3). Furthermore, these observed individual transactions are representative of the general trading pattern in the market. We verify that the total trading volume in Bitcoin by these investors correlates with the total dollar volume in the Bitcoin market, with a correlation coefficient of 0.46.

To establish the link between social learning and individual trading, we use social sentiment from conversations on Bitcointalk to predict individuals' Bitcoin transactions. Specifically, we run the following regression:

$$\text{Net Buy Dummy}_{i,t+1} = \beta \text{I(Social Sentiment)}_{i,t} + \gamma \text{Control}_{i,t} + \text{Fixed Effects} + u_{t+1}. \tag{4}$$

Given the relatively small number of users with observed Bitcoin transactions, to increase the power of the above test, we no longer restrict to users with consecutive posts (e.g., post[0] and post[1]) within a 24-hour window. Accordingly, we make a slight adjustment to the calculation of social sentiment. At each time t, we focus on users who have made at least one post over the past 7 days. Otherwise, the user-day pairs are excluded from the regression analysis above. We assume that as individuals participated on Bitcointalk, they were exposed to (and potentially affected by) other users' subsequent posts within a 24-hour window in the same thread. We analyze all the threads in which an individual has participated over the past 7 days and define social sentiment as the average sentiment of other users' posts (made within the 24-hour window following the individual's post in the same thread) during the past 7 days.

The key explanatory variable, I(Social Sentiment)$_{i,t}$, is an indicator that takes a value of one if the social sentiment encountered by individual i as of day t as defined above is positive, and zero otherwise. The dependent variable, Net Buy Dummy$_{i,t+1}$, is an indicator of whether individual i net buys Bitcoin on the next day ($t + 1$). Similar to regression (1), we include control variables such as Bitcoin return and volatility, forum sentiment, and RavenPack news sentiment, computed using data over the past 7 days. We also include user fixed effects and the prior fixed effect. We cluster standard errors at the user and date levels.

Table 12 presents our findings. As shown in column (1), when social sentiment is positive, investors are 0.371% more likely to net buy the following day. This effect size is economically significant, considering that the average probability of net buying over the next day is only 0.933%. The effect is robust when we further include control variables and date fixed effects in columns (2) and (3). Overall, social learning is not a sideshow, as social sentiment significantly predicts individuals' trading in Bitcoin.

A growing body of literature investigates the relationship between investor beliefs and their trading decisions. In a pioneering study, Giglio et al. (2021) examine the impact of belief changes on individual transactions. They establish that changes in expectations influence the direction of trading in the stock market, conditioning on occurrence of trades. Motivated by this research, we conduct a related test where we focus on consecutive posts by users (so that we can measure a user's belief change by the difference in the sentiment of post[0] and post[1]) and also require that the users trade Bitcoins around social interactions on Bitcointalk. Although the restriction that trading occurs substantially reduces the sample size for this test, results in the Appendix Table A8 show that investors' belief changes are significantly and positively related to their trading decisions.

Consistent with the findings of Armona et al. (2019) and Giglio et al. (2021), we also provide supportive evidence for the connection between investor beliefs and trading decisions, with a particular focus on belief change resulting from social learning and peer influence. Using a unique "information experiment" embedded in an online survey, Armona et al. (2019) identify the causal role of information in belief formation, as well as the causal role of beliefs in trading decisions. Our study reveals that social learning, acting as a catalyst for changes in beliefs, significantly predicts the trading direction. While Giglio et al. (2021) examine belief changes every two months, our time-stamped observations (both in posts and trading) allow us to provide confirmatory evidence at a higher frequency. Giglio et al. (2021) focus on wealthy equity investors who make cautious decisions. We extend their conclusions to speculators in the cryptocurrency market. Overall, our results highlight the role of social learning in connecting investor beliefs to trading behaviors.

5.1 Effect of Heterogeneous Social Sentiment

We present more nuanced evidence regarding the types of users that have a greater influence on others' trading decisions. Following the specifications outlined in section 4.1, we investigate three pairs of users: naive and legendary, central and non-central, and informed and non-informed users. We then repeat our analysis in regression 4, focusing specifically on the social sentiment published by distinct user groups. For example, let's consider an investor, denoted as i, who actively participated in three topics in the past seven days. In each of these topics, five different users posted within a 24-hour window before investor i's own posts. We group these 15 users based on their types and separately analyze the corresponding social sentiment for each group.

Our findings demonstrate that users' trading decisions are particularly influenced by the social sentiment of specific user groups. Table 13 presents our results. First, as indicated in columns (1) and (2), users' trading decisions respond to the social sentiment of naive users, while showing no responsiveness to the social sentiment of legendary users. Second, users' trading decisions display higher sensitivity to the social sentiment expressed by central users, as demonstrated in columns (3) and (4). Third, users' trading decisions are influenced by both informative and non-informative users' social sentiment.

Furthermore, these transaction-based findings serve to validate our primary observations on belief updating. If there exists a strong correlation between belief updates and trading, then the group of users relevant for others' belief updating should significantly overlap with users who exert influence on others' trading decisions. Indeed, we find this to be the case: the response patterns in trading decisions align well with the previously documented belief updating patterns presented in Panel B of Table 9. For example, users' belief updating is responsive to the social sentiment expressed by naive users and unresponsive to the social sentiment expressed by legendary users. We expect that their trading decisions would follow a similar pattern, and our analysis confirms this observation, as indicated in columns (1) and (2) of Table 13.

6 Social Learning and Market Outcomes

The link between social learning and trading decisions in section 5 provides a foundation for studying the effect of social learning at the aggregate level. This section connects social learning to market outcomes. To this end, we propose a daily Sentiment Contagion Intensity (SCI) index to measure the intensity of sentiment contagion at the aggregate level. As it is constructed from social interactions, it enables us to investigate market dynamics from a social perspective. We demonstrate that SCI contains novel information about future trading volume, volatility, bubbles and crashes in the Bitcoin market.

Importantly, the results in this section also effectively address concerns related to potential selection issues within our sample. If the documented social learning applies only to a small fraction of special investors in the Bitcoin market, we should not expect a substantial impact on the overall market. In other words, as long as we observe a strong aggregate impact of social learning, it appears that the selection issue in our sample may not be a significant concern.

6.1 Sentiment Contagion Intensity Index

In contrast to existing sentiment indices (such as the one proposed in Baker and Wurgler (2006)), we construct the SCI index using a bottom-up approach. First, we identify users whose sentiment changes in the direction of social sentiment after participating in a conversation on Bitcointalk. The total number of such infected users within a given day can measure the intensity of social learning. To control for the growing number of participants on Bitcointalk, we remove the time trend and seasonality by regressing the number of infected users (in logarithm) against weekday and year-month indicators. The SCI index is obtained as the residual of this regression further normalized to have a mean of zero and a standard deviation of one.

Previous studies on investor sentiment using social media data typically focus on the average sentiment level (see, e.g., Antweiler and Frank (2004)). In contrast, our SCI indicator measures the spread intensity of sentiment, making it more pertinent to the dynamics of sentiment contagion among the population. Thus, we anticipate that SCI contains information beyond the average sentiment level. The contemporaneous correlation between SCI and the average sentiment level in the forum is only 0.6%, which is also statistically insignificant. We also find that the information content of SCI has little overlap with the average sentiment level in the news media computed from the RavenPack database of Bitcoin-related news each day. The correlation coefficient between SCI and RavenPack News sentiment is -0.3%, indicating that SCI contains distinct information that is not reflected in the news sentiment. This is perhaps not surprising, given that SCI is constructed from social media, a distinct source from news media, designed to capture the amount of sentiment propagation among investors, instead of the sentiment level.

We also examine how the SCI indicator relates to several proxies for investor attention. The first proxy of investor attention is Google search volume (as explained in section 2.4). It has a correlation of 23.8% with SCI. The second proxy is based on RavenPack news coverage, measured as the number of articles on Bitcoin on each day. It has a very low correlation with SCI at 0.4%. Overall, our SCI index appears to encompass unique and novel information when compared to existing indicators, such as proxies for investor sentiment and attention.

6.2 Social Learning and Future Volume

For individual investors, as previously demonstrated in section 5, social learning plays a crucial role in predicting their trading decisions. Thus we expect a corresponding link at the aggregate level: the SCI index may serve as a predictor for trading volume in the Bitcoin market.

To examine our hypothesis, we conduct a regression analysis using the following equation:

$$\text{Ab Volume}_{t+N} = \beta_0 + \beta_1 SCI_{i,t} + \Sigma_m \gamma_m \text{Controls}_{t,m} + u_{t+k} \tag{5}$$

where Ab Volume_{t+N} denotes the abnormal dollar trading volume of Bitcoin on the next N days. Specifically, we normalize the average dollar trading volume (in billions) over the next N days ($N = 1, 7$) by subtracting its average in the past 14 days. The key predictor is $SCI_{i,t}$. As for the control variables, $\text{Controls}_{t,m}$

encompass events occurring within the last m ($m = 14$) days. To proxy for news arrivals, we calculate the average sentiment levels in Ravenpack news during the previous 14 days. Additionally, we control for Bitcoin volatility and the number of transactions by considering their cumulative sums within the past 14 days. To account for forum sentiment, we calculate the average sentiment of posts on the Bitcointalk forum published within the previous 14 days.

The findings presented in columns (1) and (2) of Table 14 strongly support our hypothesis. Overall, we observe a significant association between an increase in SCI and abnormal dollar trading volume in the subsequent days. For instance, in column (1), a one standard deviation increase in SCI leads, on average, to an additional \$4.47 million abnormal trading volume the following day. Importantly, even after controlling for news shocks, market fluctuations, and forum sentiment over the past 14 days, this relationship between SCI and abnormal trading volume remains unaffected. In column (2), we report the predictive power of SCI for the average abnormal trading volume over the next 7 days. We find that a one standard deviation increase in SCI results in an additional \$3.33 million abnormal trading volume per day over the next 7 days.

6.3 Social Learning and Future Volatility

Extensive research (Black (1986); De Long et al. (1990)) has long been devoted to examining the connection between user sentiment and asset price volatility. If users base their trading decisions on sentiment, then changes in sentiment lead to more noise trading and generate excessive volatility. Notably, recent studies (Da et al. (2014); Antweiler and Frank (2004)) have provided empirical evidence supporting this relationship, utilizing data from the U.S. stock market. However, few studies have emphasized this connection from a social perspective. Our SCI indicator naturally helps fill the gap. As highlighted in Section 4, there is a strong non-fully rational component in social learning, which results in the spread of sentiment. Consequently, as the SCI increases, social learning would trigger more subsequent noise trading and push up future volatility. Therefore, we hypothesize that the SCI index positively predicts future market volatility.

To examine this conjecture, we conduct the following regression analysis:

$$rv_{t+N} = \beta_0 + \beta_1 SCI_t + \Sigma_m \gamma_m \text{Control}_{t,m} + u_{t+N} \tag{6}$$

rv_{t+N} represents the average level of realized volatility within the next N ($N = 1, 7$) days. We compute the realized volatility by using the standard deviation of hourly returns within that day. The key explanatory variable of interest and the control variables are identical to those in equation (5).

The results of our study are presented in columns (3) and (4) of Table 14. We observe the SCI indicator positively predicts future volatility in the Bitcoin market. In column (3), a one standard deviation increase in SCI corresponds to a 0.530% increase in realized volatility over the next day. This increase amounts to 15.160% (calculated as 0.542%/3.496%) of the standard deviation of the realized volatility. In column (4), we examine how SCI predicts the average level of realized volatility over the next 7 days and find a similar pattern. Overall, results in columns (3) and (4) are consistent with our conjecture: as social interactions trigger more sentiment changes, there is an increase in noise trading by users, which subsequently raises the volatility of returns in the market.

6.4 Social Learning and Bubbles

Bubbles have long been recognized as episodes featuring elevated investor sentiment and irrational exuberance, along with a rapid rise in both asset prices and trading volume. Researchers are fascinated by these features and the underlying mechanisms behind them. The Bitcoin market provides a valuable opportunity to investigate bubbles: in the past decades, Bitcoin prices have experienced several drastic changes not easily explainable by fundamentals (Makarov and Schoar (2019)). Therefore, our data allows us to shed light on some important features of bubbles from a social learning perspective. We present an explanation from a social perspective regarding the high trading volume during bubbles.

6.4.1 Bitcoin Bubbles

To identify bubble episodes, we first label each trading day in our sample as one if the cumulative Bitcoin return over the past 90 days exceeds 200%, and zero otherwise. Next, we focus exclusively on sequences of consecutive days with a label of one that last for more than 2 quarters.[12] For each such a sequence of days, we identify the peak day as the day when the cumulative returns achieved the maximum. Finally, we designate the period between the start day and peak day as a bubble episode.

Our choice of 200% return conforms to the notion (e.g., Fama) that a bubble is associated with a

[12] We also conducted robustness checks with alternative windows, and our results remained qualitatively the same.

substantial price run-up. A return threshold of 200% effectively captures the majority of episodes that anecdotal evidence suggests were Bitcoin bubbles, including the episode towards the end of 2017. It is worth noting that we define bubbles ex-post and we do not aim to predict bubbles.

We identify four bubble episodes within our sample horizon. The first one starts on June 4, 2013, and extends to December 4, 2013. The second one begins on June 16, 2017, and ends on December 16, 2017. The third one spans from December 26, 2018, to June 26, 2019. The fourth one starts on September 13, 2020, and continues until March 13, 2021.

In Table 15, we present some key features of the identified Bitcoin bubble episodes and compare them to the non-bubble episodes. The differences in features between the bubble and non-bubble episodes are statistically significant, as supported by the joint F-statistic of 15.49 obtained from seemingly unrelated regression analysis.[13] Panel A shows that, in the identified bubble episodes, the daily average Bitcoin return is more than 8 times higher (calculated as 4.050/0.491) compared to the non-bubble episodes, and the daily return volatility is larger by 8.108% (calculated as 0.040/0.037 - 1). The total dollar volume nearly doubles during the bubble episodes. When considering news reports, the RavenPack news sentiment rises from 0.040 to 0.168, which is more than four times higher than the sentiment level in the non-bubble episodes. Moreover, Google search volume surges from 0.017 to 0.085. To summarize, the identified episodes exhibit rapid increases in returns, volume, and market fluctuations. Furthermore, there is a noticeable prevalence of optimism in media coverage, along with a surge in investor attention. All these characteristics align well with the commonly described attributes of a bubble (Shiller (2001)).

6.4.2 Social Learning and High Trading Volume in Bubbles

Why do investors trade so frequently during bubbles? Explaining the heightened trading volume presents an intriguing challenge for researchers.[14] In this section, we offer a novel perspective based on social learning, specifically the intensified contagion of optimism during bubbles. The intuition builds upon our previous findings that sentiment contagion in social interactions prompts investors to trade, subsequently predicting the trading volume at the aggregate market level. Since the contagion of optimism intensifies

[13] See, e.g., Greenwood et al. (2019) who apply seemingly unrelated regression to test the joint significance of bubble features.

[14] See Barberis et al. (2018), DeFusco et al. (2017), Liao et al. (2021).

during bubble episodes, we expect a surge in trading volume during these periods.

Social Interactions during Bubbles Sharp changes in market conditions would plausibly impact conversation patterns in various ways, including shifts in content and attitudes. In Panel B of Table 15, we summarize the crucial aspects of investor sentiment and social interactive activities during bubble episodes.

When it comes to investor sentiment, it is unsurprising to observe that users tend to be more optimistic during bubble episodes. The user sentiment increases by 12.132% (calculated as 0.305/0.272 - 1), indicating a pervasive sense of optimism in social sentiment. At the same time, user sentiment becomes less dispersed in bubbles—the standard deviation of sentiment drops by 2.232%. Such a drop in disagreement is remarkable, especially given the 14.831% (calculated as 133.504/116.261 - 1) increase in the total number of users participating in the conversations during bubbles.

In terms of social interactions, there is a significant increase in the total number of posts per day by 7.428% (calculated as 188.764/175.713 - 1). This indicates a heightened intensity in social interactions during bubbles. It is worth noting that the number of posts with positive sentiment increases disproportionately: the fraction of posts with positive sentiment increases by 5.233%. This asymmetric pattern is potentially connected to the self-enhancing transmission bias (Han et al. (2022)).[15] We provide supportive evidence in Table A9 and A10. Moreover, we observe a significant decline of 16.156% in the fraction of sophisticated users during bubbles (calculated as 1 - 0.301/0.359). This suggests that a disproportionately large number of novice investors participate in conversations during bubble episodes. These identified features of social interactions in Panel B generally align with the descriptions of bubble episodes in Shiller (2001).

Elevated Contagion of Optimism The salient features of social interactions may further reshape social learning in bubbles. As evidenced in Panel B of Table 15, the change in the fraction of positive and negative sentiment is asymmetric. To emphasize the difference in sentiment contagion between optimism and pessimism during bubbles, we present two distinct measures in Panel C of Table 15. Specifically, we identify users who become more optimistic after encountering positive social sentiment as *positively infected users*. Similarly, we define *negatively infected users*. We use the number of positively and negatively infected

[15] Days in bubbles tend to exhibit high past returns, which in turn increases the likelihood of investors publishing posts with positive sentiment.

users each day to investigate social learning in two directions during bubbles.

Our findings reveal a robust propagation of optimism and a diminished propagation of pessimism during bubbles. Compared to a day in non-bubble episodes, the number of positively infected users significantly increases by 8.927% (calculated as 16.704/15.335 - 1) in bubbles. In contrast, the number of negatively infected users on average experiences a significant decrease of 15.998% (calculated as 0.61/3.813). Collectively, the overall propagation of optimism shows a notable increase of 24.925% within one day during bubble episodes.

The notable features of social interactions in bubbles (Panel B of Table 15) provide valuable insights into the pattern. Firstly, users would frequently run into conversations with an overall positive sentiment in bubbles, as social sentiment in these periods exhibits pervasive optimism. Secondly, during bubbles, there is an increasing participation of naive investors in discussions. As we have discussed in section 4.1, these investors are more susceptible to social sentiment. Additionally, sentiment contagion becomes stronger during periods of high uncertainty. The heightened volatility in bubble episodes further amplifies the spread of optimism.

Propagation of Optimism and Trading Volume in Bubbles To effectively demonstrate the strong correlation between the propagation of optimism and trading volume in Bitcoin bubbles, a plot serves as the most suitable method. In Figure 5, we observe the temporal trend for both the number of positively infected investors (indicated by the red dashed line) on day t, and the trading volume on day $t+1$ (depicted by the blue bars). The trading volume is measured in millions of U.S. dollars. Furthermore, we provide an illustration of the Bitcoin price dynamics (represented by the green solid line) in the identified bubble episode.

We identify four Bitcoin bubble episodes spanning from 2013 to 2021, but for brevity, we will focus on the earliest episode (episode 1) and the most recent episode (episode 4).[16] In both episodes, we observe a synchronous pattern between the number of positively infected users and the trading volume on the following day. In the upper panel, we plot the bubble formation episode from January 2013 to July 2013. When the number of positively infected investors increases, the trading volume on the next day tends to increase sharply. The correlation coefficient between the number of positively infected users and trading volume is statistically

[16] The patterns in other identified episodes are similar. See Figure A1.

significant at 0.669 (p-value = 0.00). The remarkably high correlation in episode 1 becomes reasonable once we consider the dominant representativeness of the Bitcointalk forum in the early 2010s. During that time, Bitcointalk was one of the few platforms where Bitcoin investors could communicate with each other. In the lower panel, we focus on the most recent bubble-formation episode from September 2020 to May 2021. The correlation coefficient is still statistically positive, but the magnitude drops to 0.335 (p-value =0.00). This may reflect the fact that while Bitcointalk still holds significant influence, it has faced competition from other emerging social network platforms. As Bitcoin investors continue to participate in a more diverse range of social network platforms, the representativeness of Bitcointalk is inevitably decreasing.

In summary, Figure 5 presents compelling evidence illustrating the significant influence of social interactions on high trading volumes during Bitcoin bubbles, particularly through the propagation of optimism. Our results not only offer empirical validation for well-established hypotheses regarding bubbles (Shiller (2001)), but also correspond with contemporary efforts to link social dynamics with bubble phenomena (Hirshleifer (2020); Burnside et al. (2016)).

6.5 Social Learning and Market Crashes

Social learning inherently connects to price dynamics: the percolation of sentiment through social learning could influence investor demand and impose significant price pressure on the market. Moreover, with its salient nature, such as selectiveness (discussed in section 4.4), social learning can lead to the systematic proliferation of extreme sentiment. The propagation of optimism during the bubble episode is one example. This accumulated optimism may destabilize the market by initially driving asset prices too high, eventually causing a drastic correction as sentiment reverts back to normal. Consequently, we anticipate that social learning can offer a social explanation for price dynamics such as price decline and market crashes.[17]

Investors can become either more optimistic or pessimistic after social learning, leading to different directions of pressure on the price of Bitcoin. Therefore, to measure the overall price pressure associated

[17] In the presidential address, Hirshleifer (2020) also highlights the error-prone feature corresponding to the impact of social activities on the market and attributes a large part of it to the selective dissemination of sentiment, namely the social transmission bias.

with social learning, we define the Net SCI indicator as follows:

$$NetSCI_t = PosSCI_t - NegSCI_t. \tag{7}$$

The $NetSCI_t$ indicator quantifies the difference within day t in the numbers of users who are positively infected (denoted as $PosSCI_t$) and negatively infected (denoted as $NegSCI_t$). The definitions of positively infected and negatively infected users follow section 6.4.2. To capture the cumulative impact of social learning, we further calculate the sum of the Net SCI over the past 14 days. Finally, we normalize the cumulative Net SCI indicator to have a mean of zero and a standard deviation of one. We examine our hypothesis by running a Probit regression, using the cumulative net SCI as the key explanatory variable:

$$Pr[\text{Price Dynamics}_{t+N} = 1] = \Phi[\beta_0 + \beta_1 \text{Cumulative Net SCI}_t + \gamma_m \text{Controls}_{t,m} + u_{t+N}]. \tag{8}$$

We use two distinct variables to measure price dynamics. The first indicator is a return dummy, taking a value of one if the cumulative return falls below zero within a specific future horizon, and zero otherwise. This return dummy captures the price declines in the Bitcoin market. Columns (1) to (3) of Table 16 present our findings, and we report the marginal effects. An increase in the cumulative Net SCI indicator significantly elevates the probability of subsequent price declines in the Bitcoin market. Specifically, taking column (1) as an example, the interpretation of the reported marginal effects is as follows: a one standard deviation increase in cumulative Net SCI leads to a 1.813% increase in the probability of the price falling the next day in the Bitcoin market. Furthermore, as time horizons expand, the probability of further price declines also increases. For instance, in column (3), a one standard deviation increase in cumulative Net SCI raises the probability of the price falling within a 14-day window by approximately 6.228%, which is about three times higher than the probability within the next day. Overall, our findings confirm that cumulative net SCI significantly predicts future price declines.

We then zoom into the tails and link social learning to market crashes. For the dependent variable in the Probit regression (equation 8), we construct a future crash indicator that takes a value of 1 if at least one daily return falls below the 5th percentile in the future N days ($N = 1, 7, 14$). The key explanatory variable and control variables remain the same as above. Columns (4) to (6) of Table 16 demonstrate a significant relationship between increases in the SCI indicator and the probability of crashes in the Bitcoin market, as indicated by the reported marginal effects. A one standard deviation increase in the cumulative Net SCI today is associated with a statistically significant 1.038% rise in the crash probability for the following day in the Bitcoin market. This one standard deviation increase corresponds to an approximate 20.76% increase in the probability of a crash event (calculated as 1.038%/5%). As the predictive window expands from 1 day to 7 days and 14 days (columns (2) and (3)), the elevation in crash probability further amplifies to 5.696% and 9.157%, respectively. Overall, social learning strongly predicts future crashes in the Bitcoin market.

7 Conclusion

Drawing on the calls for "transition from behavioral finance to social finance" (Hirshleifer (2020)) and to gain a better understanding of "the epidemiology of narratives" (Shiller (2017)), this paper presents direct evidence for social learning. Utilizing textual analysis to extract investor sentiment from posts on Bitcointalk, a prominent online investment platform, we establish a robust channel for investors' sentiment on Bitcoin to spread through conversations.

We demonstrate that social learning is non fully rational in our context: sentiment expressed in conversations does not predict Bitcoin returns, yet investors respond to them and become more optimistic (pessimistic) about Bitcoin following social interactions on Bitcointalk that are on average positive (negative). We find evidence of both selective participation of social interactions consistent with the echo chamber effect and selective interpretation of social signals in investors' belief updating consistent with confirmation bias. Naive, less central and less informed investors are more susceptible to the influence of peers' sentiment in social interactions.

Moreover, we find that social interactions exert a substantial impact on individual investors' trading decisions as well as market outcomes. Consistent with Giglio et al. (2021), we provide supportive evidence for the connection between investor beliefs and trading decisions. Using individual trading records for a subsample of Bitcointalk users, we document that the direction of investors' trading in Bitcoin is significantly and positively related to the peer sentiment in social interactions. At the market level, the intensity of daily sentiment contagion significantly predicts Bitcoin volume and return volatility. Additionally, we develop measures of optimism derived from social interactions to assess the error-prone nature of social activities. Our results indicate that socially constructed optimism measures are significantly and positively related to the probability of future crashes of Bitcoin.

We also contribute novel insights on bubbles from the perspective of sentiment contagion. During the bubble formation episodes, we observe a heightened contagion of optimism and a larger wave of new investors engaging in social interactions. The contagion of optimism accounts for a significant portion of the variations in trading volume during bubbles. Therefore, social learning provides a fresh perspective for understanding the elevated trading volume during bubbles.

This paper strongly advocates for the significance of social finance in comprehending investors' decision-making processes and market dynamics. The documented phenomenon of sentiment contagion through conversations not only provides fresh perspectives on investors belief dynamics but also underscores the potential of social interactions in elucidating various financial phenomena. The online investment community serves as an ideal platform for investigating other pertinent questions in the realm of social finance, and we leave these inquiries to future research endeavors.

References

Altonji, J. G., Elder, T. E., and Taber, C. R. (2005). Selection on observed and unobserved variables: Assessing the effectiveness of catholic schools. *Journal of political economy*, 113(1):151–184.

Antweiler, W. and Frank, M. Z. (2004). Is all that talk just noise? the information content of internet stock message boards. *The Journal of Finance*, 59(3):1259–1294.

Armona, L., Fuster, A., and Zafar, B. (2019). Home price expectations and behaviour: Evidence from a randomized information experiment. *The Review of Economic Studies*, 86(4):1371–1410.

Baker, M. and Wurgler, J. (2006). Investor sentiment and the cross-section of stock returns. *The journal of Finance*, 61(4):1645–1680.

Baker, S. R., Bloom, N., and Davis, S. J. (2016). Measuring economic policy uncertainty. *The Quarterly Journal of Economics*, 131(4):1593–1636.

Barber, B. and Odean, T. (2013). The behavior of individual investors. volume 2, chapter Chapter 22, pages 1533–1570. Elsevier.

Barber, B. M. and Odean, T. (2000). Trading is hazardous to your wealth: The common stock investment performance of individual investors. *The journal of Finance*, 55(2):773–806.

Barberis, N., Greenwood, R., Jin, L., and Shleifer, A. (2018). Extrapolation and bubbles. *Journal of Financial Economics*, 129(2):203–227.

Bénabou, R. and Tirole, J. (2016). Mindful economics: The production, consumption, and value of beliefs. *Journal of Economic Perspectives*, 30(3):141–164.

Black, F. (1986). Noise. *The Journal of Finance*, 41(3):528–543.

Brown, B., Bakken, J., Ameringer, S., and Mahon, S. (2008). A comprehensive conceptualization of the peer influence process in adolescence. *Understanding Peer Influence In Children and Adolescents*, pages 17–44.

Brunnermeier, M. K. and Parker, J. A. (2005). Optimal expectations. *American Economic Review*, 95(4):1092–1118.

Burnside, C., Eichenbaum, M., and Rebelo, S. (2016). Understanding booms and busts in housing markets. *Journal of Political Economy*, 124(4):1088–1147.

Chen, H., De, P., Hu, Y., and Hwang, B.-H. (2014). Wisdom of crowds: The value of stock opinions transmitted through social media. *The review of financial studies*, 27(5):1367–1403.

Cong, L. W., Li, Y., and Wang, N. (2021). Tokenomics: Dynamic adoption and valuation. *The Review of Financial Studies*, 34(3):1105–1155.

Cookson, J. A., Engelberg, J. E., and Mullins, W. (2023). Echo chambers. *The Review of Financial Studies*, 36(2):450–500.

Da, Z., Engelberg, J., and Gao, P. (2011). In search of attention. *The Journal of Finance*, 66(5):1461–1499.

Da, Z., Engelberg, J., and Gao, P. (2014). The sum of all fears investor sentiment and asset prices. *The Review of Financial Studies*, 28(1):1–32.

De Long, J. B., Shleifer, A., Summers, L. H., and Waldmann, R. J. (1990). Noise trader risk in financial markets. *Journal of Political Economy*, 98(4):703–738.

DeFusco, A. A., Nathanson, C. G., and Zwick, E. (2017). Speculative dynamics of prices and volume. Technical report, National Bureau of Economic Research.

Faia, E., Fuster, A., Pezone, V., and Zafar, B. (2022). Biases in information selection and processing: Survey evidence from the pandemic. *Review of Economics and Statistics*, pages 1–46.

Feng, L. and Seasholes, M. S. (2004). Correlated trading and location. *The Journal of Finance*, 59(5):2117–2144.

Giglio, S., Maggiori, M., Stroebel, J., and Utkus, S. (2021). Five facts about beliefs and portfolios. *American Economic Review*, 111(5):1481–1522.

Greenwood, R., Shleifer, A., and You, Y. (2019). Bubbles for fama. *Journal of Financial Economics*, 131(1):20–43.

Griffin, J. M. and Shams, A. (2020). Is bitcoin really untethered? *The Journal of Finance*, 75(4):1913–1964.

Hagberg, A., Swart, P., and S Chult, D. (2008). Exploring network structure, dynamics, and function using networkx. Technical report, Los Alamos National Lab.(LANL), Los Alamos, NM (United States).

Han, B., Hirshleifer, D., and Walden, J. (2022). Social transmission bias and investor behavior. *Journal of Financial and Quantitative Analysis*, 57(1):390–412.

Heimer, R. Z. (2016). Peer pressure: Social interaction and the disposition effect. *The Review of Financial Studies*, 29(11):3177–3209.

Hirshleifer, D. (2020). Presidential address: Social transmission bias in economics and finance. *The Journal of Finance*, 75(4):1779–1831.

Hodrick, R. J. (1992). Dividend yields and expected stock returns: Alternative procedures for inference and measurement. *The Review of Financial Studies*, 5(3):357–386.

Hong, H., Kubik, J. D., and Stein, J. C. (2004). Social interaction and stock-market participation. *The Journal of Finance*, 59(1):137–163.

Hong, H., Kubrik, J. D., and Stein, J. C. (2005). Thy neighbor's portfolio: Word-of-mouth effects in the holdings and trades of money managers. *The Journal of Finance*, 60(6):2801–2824.

Huang, S., Hwang, B.-H., and Lou, D. (2021). The rate of communication. *Journal of Financial Economics*, 141(2):533–550.

Hwang, B.-H. (2023). The impact of word-of-mouth communication on investors' decisions and asset prices. In *Handbook of Financial Decision Making*, pages 171–191. Edward Elgar Publishing.

Jiang, W. (2017). Have instrumental variables brought us closer to the truth. *Review of Corporate Finance Studies*, 6(2):127–140.

Kogan, S., Makarov, I., Niessner, M., and Schoar, A. (2024). Are cryptos different? evidence from retail trading. *Journal of Financial Economics*. Forthcoming.

Kuchler, T., Li, Y., Peng, L., Stroebel, J., and Zhou, D. (2022). Social proximity to capital: Implications for investors and firms. *The Review of Financial Studies*, 35(6):2743–2789.

Lee, C. M. C., Shleifer, A., and Thaler, R. H. (1991). Investor sentiment and the closed-end fund puzzle. *Journal of Finance*, 46(1):75–109.

Liao, J., Peng, C., and Zhu, N. (2021). Extrapolative bubbles and trading volume. *Review of Financial Studies*.

Liu, Y., Sheng, J., and Wang, W. (2022a). Technology and cryptocurrency valuation: Evidence from machine learning. *Working Paper*.

Liu, Y. and Tsyvinski, A. (2021). Risks and returns of cryptocurrency. *The Review of Financial Studies*, 34(6):2689–2727.

Liu, Y., Tsyvinski, A., and Wu, X. (2022b). Common risk factors in cryptocurrency. *The Journal of Finance*, 77(2):1133–1177.

Makarov, I. and Schoar, A. (2019). Price discovery in cryptocurrency markets. In *AEA Papers and Proceedings*, volume 109, pages 97–99. American Economic Association.

Makarov, I. and Schoar, A. (2020). Trading and arbitrage in cryptocurrency markets. *Journal of Financial Economics*, 135(2):293–319.

Manning, C. D., Surdeanu, M., Bauer, J., Finkel, J. R., Bethard, S., and McClosky, D. (2014). The stanford corenlp natural language processing toolkit. In *Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations*, pages 55–60.

Mian, A. and Sufi, A. (2014). What explains the 2007–2009 drop in employment? *Econometrica*, 82(6):2197–2223.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of general psychology*, 2(2):175–220.

Oster, E. (2019). Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics*, 37(2):187–204.

Shiller, R. J. (2001). *Irrational Exuberance*. Broadway Books.

Shiller, R. J. (2007). Understanding recent trends in house prices and home ownership.

Shiller, R. J. (2017). Narrative economics. *American economic review*, 107(4):967–1004.

Shiller, R. J., Fischer, S., and Friedman, B. M. (1984). Stock prices and social dynamics. *Brookings papers on economic activity*, 1984(2):457–510.

Shiller, R. J. and Pound, J. (1989). Survey evidence on diffusion of interest and information among investors. *Journal of Economic Behavior & Organization*, 12(1):47–66.

Simon, D. and Heimer, R. (2012). Facebook finance: How social interaction propagates active investing. In *AFA 2013 San Diego Meetings Paper*.

Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance*, 62(3):1139–1168.

Tversky, A. and Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157):1124–1131.

Figure 1: Example of Conversation and Illustration of Empirical Strategy

This figure presents an example for a conversation in our sample and our empirical strategy. We measure sentiment changes by calculating the difference between sentiment in post[0] and post[1]. We measure social sentiment by calculating the average sentiment of others' post in conversations.

Figure 2: Most Common Words in Posts

This word cloud shows the most frequent words present in posts. The size of the words depends on their relative frequency, hence more common words are larger.

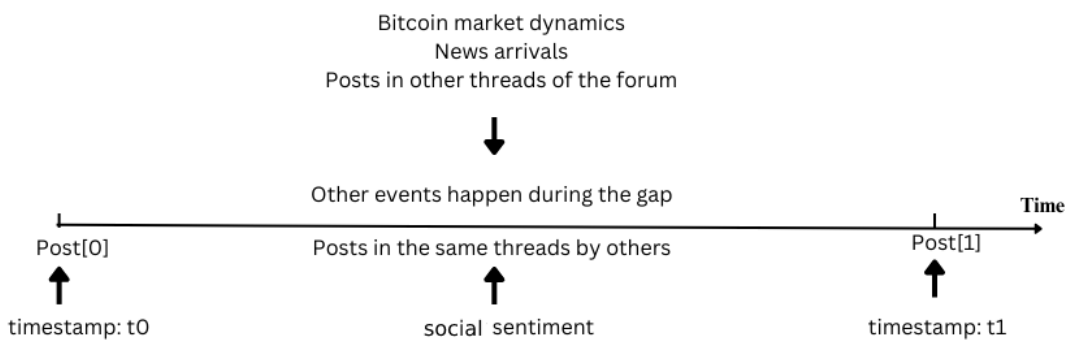


Figure 3: Timeline and Illustration of Empirical Strategy

This figure presents the timeline of post[0], post[1] and social sentiment. It also provides a formal demonstration of our empirical strategy.

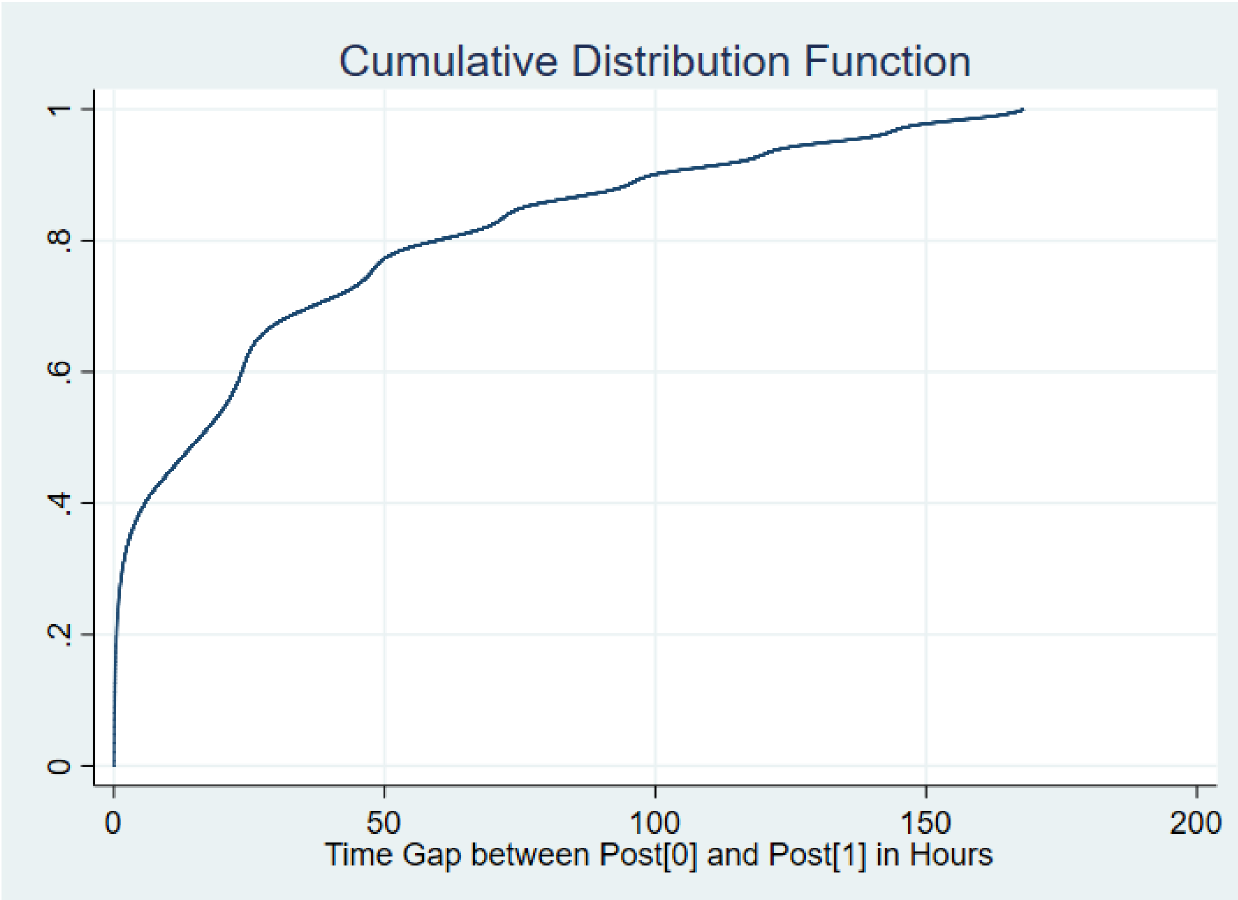


Figure 4: Cumulative Distribution Function of Time Gap

This figure presents the cumulative distribution function of the time gap between two consecutive posts by the same user.

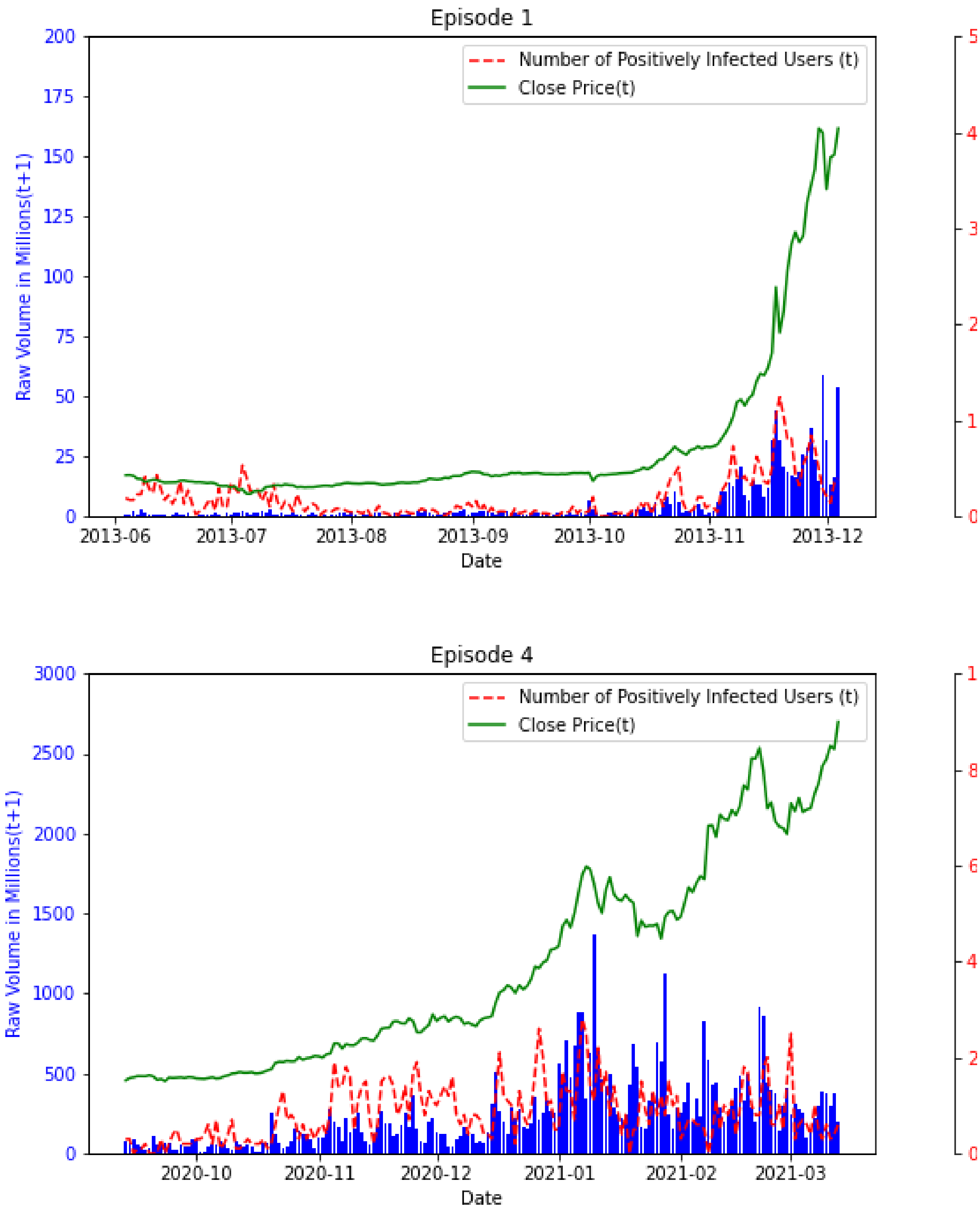


Figure 5: Sentiment Contagion and Trading Volume in Bubble Episodes

This figure plots the number of positively infected users on day t (the red dash line) and trading volume (the blue bars; in millions) on day $t+1$, along with the Bitcoin price index on day t (the green solid line). We illustrate the first (upper panel) and last (lower panel) identified bubble episodes in our sample.

Table 1: Summary Statistics

This table tabulates the summary statistics of the key variables. In Panel A, we present statistics of user sentiment in posts aggregated at the user level. To obtain the summary statistics for the average sentiment, we first calculate the mean sentiment of posts published by each user, then report the corresponding statistics based on the whole user population. In Panel B, we present statistics of user sentiment aggregated at the daily level. In Panel C, we present statistics of user sentiment aggregated at the thread level. In Panel D, we present summary statistics of market variables, RavenPack news sentiment and google search volume. Mean Return is the annualized average daily return of Bitcoin. Return Volatility (within Day) is the within-day volatility of Bitcoin based on the hourly Bitcoin return. Number of Transactions is the total number of Bitcoin transactions within one day. Number of Bitcoins Traded is the total number of Bitcoins being traded within one day. Total Dollar Volume (in millions) is the total dollar trading volume of Bitcoin measured in millions of dollars within one day. RavenPack News Sentiment is the sentiment score for the news on Bitcoin in the RavenPack database. We normalize the sentiment score to be between $[-1, 1]$. Google Search (Bitcoin) records the detrended Google Search Index for Bitcoin on a daily basis. Our sample spans from May 1, 2012 to July 30, 2022.

	count	mean	p50	sd	min	max
Panel A: Post Activities at User level						
Average Sentiment	44,356	0.324	0.333	0.512	-1.000	1.000
Standard Deviation of Sentiment	28,077	0.608	0.640	0.297	0.000	1.414
Number of Posts	44,356	15.015	2	72.469	1	3851
Panel B: Post Activities at Daily level						
Average Sentiment	3,737	0.279	0.282	0.118	-0.250	0.833
Standard Deviation of Sentiment	3,736	0.669	0.672	0.054	0.392	0.917
Number of Users	3,737	119.601	85	113.390	1	980
Number of Posts	3,737	178.219	128	160.778	1	1400
Panel C: Post Activities at Thread level						
Average Sentiment	18,799	0.237	0.250	0.324	-1.000	1.000
Standard Deviation of Sentiment	17,191	0.646	0.659	0.189	0.000	1.414
Number of Users	18,799	24.838	11	55.594	1	2589
Number of Posts	18,799	35.428	13	551.522	1	73859
Median Gap in Days between Consecutive Posts	17,214	0.573	0.033	16.568	0.000	1129.115
Panel D: Other Data Sources						
Daily Return(Annualized)	3,743	1.190	0.714	16.166	-177.093	146.518
Return Volatility(Within Day)	3,743	0.037	0.029	0.035	0.000	0.599
Number of Transactions	3,743	16471.536	11376	17268.770	0	181616
Number of Bitcoins Traded	3,743	9358.724	6660.652	9697.557	0	137070.178
Total Dollar Volume (in millions)	3,743	56.499	18.834	97.465	0.000	1372.717
RavenPack News Sentiment	3,743	0.041	0.000	0.330	-0.660	0.660
Google Search (Bitcoin)	3,743	0.031	-0.046	0.417	-0.905	6.595

Table 2: Contagion Effect

This table presents the panel regression analysis of sentiment change on social sentiment. The dependent variable is the sentiment change defined as the revision in sentiment between a user's two consecutive posts post[0] and post[1]. The main explanatory variable is the social sentiment defined as the average sentiment of others' posts within the conversation, published between the timestamps of post[0] and post[1]. We control for simultaneous events between post[0] and post[1], such as Bitcoin return and volatility, activities in other conversations within the forum, and news arrivals documented in the RavenPack database. We also include lagged news arrivals to capture a delay in response. In all columns, we include a prior fixed effect. The t-statistics (in parentheses) are based on standard errors clustered by user and day. We multiply each coefficient by 100.

	Sentiment Change				
	(1)	(2)	(3)	(4)	(5)
Social Sentiment	2.838***	2.836***	3.153***	3.096***	3.088***
	(6.81)	(6.82)	(6.05)	(6.02)	(6.01)
RavenPack News Sentiment between Post[0] and Post[1]		0.329	-0.271	-0.195	-0.032
		(0.41)	(-0.32)	(-0.23)	(-0.04)
Bitcoin Return			29.443***	29.927***	30.723***
			(5.76)	(5.80)	(5.90)
Bitcoin Volatility			15.087	13.439	12.925
			(0.95)	(0.84)	(0.81)
Forum Sentiment				-4.659***	-4.679***
				(-4.39)	(-4.42)
RavenPack News Sentiment 24 hours before Post[0]					1.385
					(1.45)
RavenPack News Sentiment 48 hours before Post[0]					-1.030
					(-1.03)
Prior FE	YES	YES	YES	YES	YES
User FE	YES	YES	YES	YES	YES
Date FE	YES	YES	YES	YES	YES
Adjusted R-Squared	0.344	0.344	0.345	0.345	0.345
N	212,623	212,623	178,614	175,750	175,750

Table 3: Placebo Test

The table displays the results of a placebo test examining the contagion effect. The dependent variable is the sentiment change defined as the revision in sentiment between a user's two consecutive posts post[0] and post[1]. The independent variable is the average sentiment level in N random conversations ($N = 1, 3, 5, 10$) that the user did not participate in, which occurred between the timestamps of post[0] and post[1]. The control variables remain identical to those in Table 2. In all columns, we include a prior fixed effect. The t-statistics (in parentheses) are based on standard errors clustered by user and day. We multiply each coefficient by 100.

	Sentiment Change			
	(1)	(2)	(3)	(4)
Random Sentiment	0.158	-0.061	-0.378	-0.079
	(0.36)	(-0.14)	(-0.77)	(-0.15)
RavenPack News Sentiment	0.216	0.356	0.291	0.903
between Post[0] and Post[1]	(0.27)	(0.42)	(0.33)	(0.96)
Bitcoin Return	30.583***	30.916***	31.107***	30.741***
	(6.36)	(6.13)	(6.41)	(6.17)
Bitcoin Volatility	9.781	8.801	4.081	11.997
	(0.71)	(0.67)	(0.29)	(0.77)
Forum Sentiment	-5.292***	-7.831***	-10.699***	-15.231***
	(-5.45)	(-6.24)	(-6.76)	(-7.19)
RavenPack News Sentiment	1.232	1.678*	1.657*	1.520
24 hours before Post[0]	(1.39)	(1.80)	(1.72)	(1.49)
RavenPack News Sentiment	-1.361	-1.006	-0.478	-0.238
48 hours before Post[0]	(-1.42)	(-1.03)	(-0.47)	(-0.22)
Prior FE	YES	YES	YES	YES
User FE	YES	YES	YES	YES
Date FE	YES	YES	YES	YES
Controls	YES	YES	YES	YES
Adjusted R-Squared	0.345	0.346	0.346	0.346
N	195,635	180,813	169,124	148,284

Table 4: Contagion Effect: Oster Test

This table tests the role of omitted and unobservable control variables in driving the relation between social sentiment and sentiment change using the test proposed in Oster (2019). To implement this, we estimate a linear model:

$$\begin{aligned}\text{Senti Change}_{i,j,t_0\to t_1} &= \beta_1 \text{Social Sentiment}_{i,j,t_0\to t_1} \\ &+ \gamma' \text{Control}_{i,t_0\to t_1} + \text{Fixed Effects} + u_{i,t_1}.\end{aligned}$$

First, we estimate the model without any control variables, from which we obtain the coefficients β_u and the coefficient of determination R_u^2. Then, we include contemporaneous and lagged control variables in our main results, from which we obtain the coefficients β_c and the coefficient of determination R_c^2 (as reported in the last column of Table 2). For any given combination of test parameters, denoted as δ and R^2_{max}, Oster (2019) defines the bias-adjusted coefficient β_{adj}. The adjusted coefficient β_{adj} is closely approximated by (or strictly equal to when δ equals one):

$$\beta_{adj} \approx \beta_c - \delta \frac{(\beta_u - \beta_c)(R^2_{max} - R^2_c)}{R^2_c - R^2_u}.$$

With this adjusted coefficient β_{adj}, the recommended identified set is the interval between β_{adj} and β_c. In this table, we report the identified set for different combinations of parameters. Additionally, we report whether the identified set rejects the null hypothesis of $\beta = 0$.

$R^2_{max} = \min(2.2R^2_c, 1)$

$\delta = 1$			$\delta = 2$		
β_{adj}	Identified Set	Reject Null?	β_{adj}	Identified Set	Reject Null?
0.083	[0.031, 0.083]	Y	0.177	[0.031, 0.177]	Y

$R^2_{max} = \min(3R^2_c, 1)$

$\delta = 1$			$\delta = 2$		
β_{adj}	Identified Set	Reject Null?	β_{adj}	Identified Set	Reject Null?
0.105	[0.031, 0.105]	Y	0.281	[0.031, 0.281]	Y

Table 5: Contagion Effect: Sentiment Change for Frequent Posters and Infreqent Posters

This table presents a subsample analysis of sentiment change on social sentiment. Specifically, we divide the user sample into two groups: frequent posters, who publish posts frequently, and infrequent posters, who publish posts infrequently. we calculate the average interval of her consecutive pairs of posts using a rolling window approach on a daily basis. We define users with an average interval above the population average as infrequent posters, and the remaining users as frequent posters. For each subsample analysis in columns (1) to (2), we adopt the regression settings and variable definitions from column (5) in Table 2. We investigate the interactive effect in column (3). The t-statistics (in parentheses) are based on standard errors clustered by user and day. We multiply each coefficient by 100.

	Infrequent Posters	Frequent Posters	Interactive Effect
	(1)	(2)	(3)
Social Sentiment	3.503**	2.970***	4.095***
	(2.57)	(5.30)	(3.34)
Frequent Poster			-1.254
* Social Sentiment			(-0.95)
Frequent Poster			0.308
			(0.33)
Prior FE	YES	YES	YES
User FE	YES	YES	NO
Date FE	YES	YES	NO
Controls	YES	YES	YES
Adjusted R-Squared	0.343	0.345	0.346
N	31,525	137,437	170,913

Table 6: Effect of Social Sentiment: Heterogeneity across Users

In this table, we explore different users' responses to social sentiment. The dependent variable is the sentiment change defined as the revision in sentiment between a user's two consecutive posts post[0] and post[1]. In column (1), we examine whether naive users, indicated by a dummy variable that equals one for users with non-lengendary status on Bitcointalk forum, are more susceptible to social sentiment. In column (2), we investigate whether users who have participated in more conversations are less influenced by social sentiment. This user feature measures a user's centrality based on past posts on the social network, where two users are considered connected if they have participated in the same thread. In column (3), we study whether less informed users, defined as those whose post sentiment has a low correlation with future Bitcoin returns in 7 days (correlation $<$ 25th percentile), are more subject to social sentiment. In all columns, we include a prior fixed effect. The coefficients are multiplied by 100, and the t-statistics (in parentheses) are based on standard errors clustered by user and day.

	Naive Users	Central Users	Less Informed Users
	(1)	(2)	(3)
User Feature * Social Sentiment	3.585***	-0.047***	2.664**
	(3.46)	(-3.82)	(1.98)
Social Sentiment	1.125	4.086***	2.105***
	(1.52)	(7.28)	(3.49)
User Feature		-0.006	-0.268
		(-0.66)	(-0.36)
Prior FE	YES	YES	YES
User FE	YES	YES	YES
Date FE	YES	YES	YES
Controls	YES	YES	YES
Adjusted R-Squared	0.345	0.345	0.343
N	175,750	175,750	139,873

Table 7: Time-variation in Sentiment Contagion

In this table, we examine sentiment contagion intensity variations across episodes. The dependent variable is the sentiment change measured as the revision in sentiment between a user's two consecutive posts post[0] and post[1]. In column (1), we investigate the impact of informative days on sentiment contagion by defining informative days based on the number of news arrivals documented in the RavenPack database. If there are no news arrivals on a specific day, we assign a value of 0. We construct a dummy variable called *informative days* that takes the value of 1 on a given day if the total number of news arrivals is higher than the sample median. In column (2), we investigate how information uncertainty affects sentiment contagion. To identify days of high information uncertainty, we use the standard deviation of news sentiment within the day documented in the RavenPack database. We construct a dummy variable called *high information uncertainty*, which takes the value of 1 if this standard deviation is higher than the sample median and 0 otherwise. If there are no news arrivals on a specific day, we assign a value of 0 to the dummy. In column (3), we examine whether Bitcoin return volatility affects sentiment contagion. To do so, we calculate the standard deviation of hourly Bitcoin return within the past 24-hour window before the first post is published. We construct a dummy variable called *high Bitcoin volatility*, which equals 1 if the standard deviation is higher than the sample mean and 0 otherwise. In all columns, we include a prior fixed effect. The coefficients are multiplied by 100, and the t-statistics (in parentheses) are based on standard errors clustered by user and day.

	Informative Days	High Uncertainty	High Bitcoin Volatility
	(1)	(2)	(3)
Episode Feature * Social Sentiment	0.959	3.851**	2.919**
	(0.92)	(2.36)	(2.50)
Social Sentiment	2.664***	3.613***	2.856***
	(3.76)	(5.42)	(3.59)
Episode Feature		-1.059	-1.501*
		(-1.10)	(-1.96)
Prior FE	YES	YES	YES
User FE	YES	YES	YES
Date FE	YES	YES	YES
Controls	YES	YES	YES
Adjusted R-Squared	0.345	0.322	0.322
N	175,750	143,092	143,092

Table 8: Social Sentiment and Future Returns

This table presents the panel regression analysis of future Bitcoin returns on social sentiment constructed from Bitcointalk posts. The dependent variable is the Bitcoin returns in future K hours (K = 6, 24, 48, 72, and 168), starting from the ending hour of post[1]. The key predictor is the conversation-level social sentiment. Control variables are identical to those in Table 2. In all columns, we include a prior fixed effect. The t-statistics (in parentheses) are based on standard errors clustered by user and day. We multiply each coefficient by 100.

	(1) 6 Hours	(2) 24 Hours	(3) 48 Hours	(4) 72 Hours	(5) 168 Hours
Social Sentiment	-0.012	-0.094**	-0.119***	-0.111**	-0.114**
	(-0.38)	(-2.17)	(-2.69)	(-2.31)	(-2.57)
RavenPack News Sentiment	-0.733***	-1.859***	-1.831***	-1.369***	-1.786***
between Post[0] and Post[1]	(-3.43)	(-5.61)	(-6.76)	(-5.42)	(-5.09)
Bitcoin Return	-21.330***	-35.052***	-39.628***	-41.422***	-41.485***
	(-13.37)	(-8.78)	(-7.81)	(-9.48)	(-10.99)
Bitcoin Volatility	17.717	15.433	45.070***	41.646***	37.695***
	(1.44)	(0.75)	(5.30)	(5.36)	(3.04)
Forum Sentiment	-0.245***	-0.537***	-0.499***	-0.420***	-0.418***
	(-3.82)	(-4.68)	(-4.51)	(-3.68)	(-3.39)
RavenPack News Sentiment	-1.566***	-2.586***	-2.778***	-2.223***	-2.624***
24 hours before Post[0]	(-4.28)	(-5.93)	(-7.26)	(-7.66)	(-6.79)
RavenPack News Sentiment	-0.735***	-1.570***	-1.030***	-0.862***	-1.524***
48 hours before Post[0]	(-2.93)	(-4.10)	(-2.77)	(-2.71)	(-4.67)
Prior FE	YES	YES	YES	YES	YES
User FE	YES	YES	YES	YES	YES
Date FE	YES	YES	YES	YES	YES
Adjusted R-Squared	0.268	0.675	0.824	0.878	0.939
N	175,750	175,750	175,750	175,750	175748

Table 9: Sentiment Contagion of Posts by Different Types of Users

This table presents how users respond to social sentiment published by different types of users with different level of informedness. We examine three user features: (i) legendary users whose ranking on Bitcointalk is labeled as "Legendary"; (ii) central users who have participated in more conversations; (iii) informed users whose post sentiment has a high correlation with future Bitcoin returns in 7 days (the correlation exceeds the 25th percentile of the population). In Panel A, we analyze how social sentiment by different types of users predict future 24-hour returns, with a similar setting in Table 8. In Panel B, we investigate how users revise their beliefs after reading social sentiment by users of different types. To account for variations, we include user and day fixed effects as well as a prior fixed effect in all columns of Panel B. Control variables are identical to those in Table 2. The t-statistics (shown in parentheses) are based on standard errors that are clustered by user and day. Additionally, we multiply each coefficient by 100 in all panels.

Panel A: Informativeness (future 24-hour returns)

	Legendary		Central		Informed	
	(1)	(2)	(3)	(4)	(5)	(6)
Social Sentiment	-0.054		-0.085**		-0.079*	
by Featured Users	(-1.12)		(-2.23)		(-1.78)	
Social Sentiment		-0.061*		-0.036		-0.179***
by NonFeatured Users		(-1.71)		(-0.99)		(-2.58)
Prior FE	YES	YES	YES	YES	YES	YES
User FE	YES	YES	YES	YES	YES	YES
Date FE	YES	YES	YES	YES	YES	YES
Controls	YES	YES	YES	YES	YES	YES
Adjusted R-Squared	0.674	0.676	0.673	0.695	0.674	0.676
N	138938	161144	159205	119231	136962	36580

Panel B: Response

	Legendary		Central		Informed	
	(1)	(2)	(3)	(4)	(5)	(6)
Social Sentiment	0.804*		2.897***		2.712***	
by Featured Users	(1.66)		(5.55)		(4.80)	
Social Sentiment		2.399***		1.925***		4.897***
by NonFeatured Users		(4.94)		(2.84)		(3.90)
Prior FE	YES	YES	YES	YES	YES	YES
User FE	YES	YES	YES	YES	YES	YES
Date FE	YES	YES	YES	YES	YES	YES
Controls	YES	YES	YES	YES	YES	YES
Adjusted R-Squared	0.347	0.345	0.345	0.346	0.344	0.338
N	138,938	161,144	173,477	43,701	136,962	36,580

Table 10: Confirmation Bias: Echo Chambers

In this table, we use the Probit model to examine how users with different priors selectively participate in conversations. We focus on consecutive pairs of posts (post[0] and then post[1]) published within a 24-hour period by the same user. Our primary explanatory variable is "positive prior", which takes a value of 1 if the user's prior sentiment (i.e., sentiment in post[0]) is positive and 0 otherwise. The dependent variable reflects the attitude of the subsequent conversation in which the user takes part (i.e., the conversation to which post[1] belongs) and is represented by a dummy variable that equals 1 if it signals positive sentiment and 0 otherwise. To determine whether the subsequent conversation signals positive sentiment, we employ two methods. In columns (1) and (2), we set the dependent variable to 1 if the thread to which post[1] belongs has a starting post with positive sentiment. In columns (3) and (4), we set the dependent variable to 1 if the social sentiment (i.e., the average sentiment of the posts between post[0] and then post[1]) is positive. The control variables are identical to those in Table 2. We report the average marginal effects. The t-statistics (in parentheses) are computed using standard errors clustered by user and day. Additionally, we multiply each coefficient by 100.

	Positive Sentiment of First Post		Positive Social Sentiment	
	(1)	(2)	(3)	(4)
Positive Prior	6.738***	5.726***	4.252***	3.534***
	(19.65)	(15.97)	(17.84)	(14.39)
Controls	No	Yes	No	Yes
Adjusted R-Squared	0.004	0.019	0.002	0.013
N	190,085	131,407	142,871	123,084

Table 11: Confirmation Bias: Selective Interpretation

This table investigates whether users' response to social sentiment varies depending on whether the sentiment aligns with their priors. To do so, we decompose social sentiment into two parts. Social sentiment (+) is the positive part of social sentiment, and Social sentiment (-) is the negative part. In the first two columns, we analyze the subset of users with positive priors and observe how they respond to positive and negative social sentiment. In column (1), the explanatory variable is the average sentiment in conversations. In column (2), the explanatory variables are two separate sentiment variables. In the last two columns, we investigate the response of users with negative priors to positive and negative social sentiment. In column (3), the explanatory variable is the average sentiment in conversations. In column (4), the explanatory variables are two separate sentiment variables. We include user and day fixed effects in all columns. We report t-statistics in parentheses based on standard errors clustered by user and day. Finally, we multiply each coefficient by 100.

	Positive Priors		NonPositive Priors	
	(1)	(2)	(3)	(4)
Social Sentiment	4.279***		3.294***	
	(6.06)		(4.68)	
Social Sentiment(+)		5.012***		2.764***
		(4.96)		(2.74)
Social Sentiment(-)		2.687		4.283***
		(1.61)		(2.78)
User FE	YES	YES	YES	YES
Date FE	YES	YES	YES	YES
Controls	YES	YES	YES	YES
Adjusted R-Squared	0.044	0.044	0.249	0.249
N	85,949	85,949	86,237	86,237

Table 12: Social Learning and Individual Trading

This table presents the regression analysis of one individual's trading decisions on the social sentiment that an investor encounters. The dependent variable is an indicator of whether the individual net buys on day $t+1$. The main explanatory variable is the social sentiment, and the definition follows section 5. Control variables are identical to those in Table 2. In all columns, we include user fixed effect and a prior fixed effect. We report t-statistics in parentheses based on standard errors clustered by user and day. We multiply each coefficient by 100.

	Probability of net buying in the next day		
	(1)	(2)	(3)
I(SocialExposure Sentiment>0)	0.371***	0.343***	0.293***
	(3.33)	(2.86)	(2.61)
Prior FE	YES	YES	YES
User FE	YES	YES	YES
Date FE	NO	NO	YES
Controls	NO	YES	NO
Adjusted R-Squared	0.081	0.081	0.096
N	249,449	249,180	249,449

Table 13: Effect of Heterogeneous Social Sentiment on Trading

This table presents the heterogeneous effect of social sentiment on trading decisions across different groups. The dependent variable is an indicator of whether individuals make net buys on day $t+1$. The main explanatory variable is the social sentiment from various user groups. Control variables are identical to those in Table 2. In all columns, we include user fixed effect and a prior fixed effect. We report t-statistics in parentheses based on standard errors clustered by user and day. We multiply each coefficient by 100.

	(1)	(2)	(3)	(4)	(5)	(6)
I(Sentiment>0)	0.175	0.156	0.161	0.174	0.124	0.146
	(1.60)	(1.45)	(1.47)	(1.64)	(1.15)	(1.34)
I (Legendary Exposure Sentiment>0)	-0.115					
	(-0.52)					
I (Naive Exposure Sentiment>0)		0.445***				
		(3.24)				
I (Central Exposure Sentiment>0)			0.419*			
			(1.92)			
I (Non-central Exposure Sentiment>0)				-0.060		
				(-0.28)		
I (Informed Exposure Sentiment>0)					0.595***	
					(4.02)	
I (Non-informed Exposure Sentiment>0)						0.811***
						(3.27)
Adjusted R-Squared	0.096	0.096	0.096	0.096	0.096	0.096
N	249,449	249,449	249,449	249,449	249,449	249,449
User FE	YES	YES	YES	YES	YES	YES
Date FE	NO	NO	NO	NO	NO	NO
Controls	YES	YES	YES	YES	YES	YES

Table 14: Social Learning, Market Volume and Market Volatility

This table presents the predictive power of social learning on the future volume and volatility in the Bitcoin market. The main explanatory variable is the SCI indicator, which measures the daily number of users who adjust their sentiment to align with the social sentiment in conversations. We eliminate the time trend by regressing the raw number series (in logarithm) on year-month indicators and weekday dummies. We obtain our SCI indicator by further normalizing the residual series to have a mean of zero and a standard deviation of one. In columns (1) and (2), the primary dependent variable is the future dollar trading volume of Bitcoin (in billions) for the upcoming 1 day and 7 days. We normalize the trading volume by subtracting its average in the past two weeks. For columns (3) and (4), the main dependent variable is the future volatility of Bitcoin returns over the next 1 day and 7 days. To account for news arrivals, we include the average levels of sentiment in RavenPack news over the past 14 days as control variables. Additionally, we incorporate two Bitcoin market variables: Bitcoin Volatility and the Number of Transactions. Each variable represents the cumulative sum over the past 14 days. To gauge sentiment in online forums, we calculate the average sentiment of posts on the Bitcointalk forum published in the previous 14 days. We report t-statistics in parentheses based on Newey-West corrected standard errors using a lag of 14 days. We multiply each coefficient by 100.

	Abnormal Trading Volume		Return Volatility	
	(1)	(2)	(3)	(4)
	1 Day	7 Days	1 Day	7 Days
SCI	0.447***	0.333***	0.530***	0.370***
	(3.85)	(3.04)	(5.45)	(4.33)
Bitcoin Volatility	-4.678***	-5.702***	11.237***	8.597***
in Past 14 Days	(-2.92)	(-3.04)	(7.60)	(8.81)
RavenPack News Sentiment	-0.291	0.273	-0.512***	-0.425***
in Past 14 Days	(-1.01)	(1.16)	(-4.41)	(-4.45)
Forum Average Sentiment	0.447	0.819	-1.555***	-1.600***
in Past 14 Days	(0.95)	(1.38)	(-3.92)	(-3.43)
Bitcoin Return	3.509***	4.059***	2.539***	3.913***
in Past 14 Days	(3.26)	(3.55)	(2.75)	(3.17)
Adjusted R-Squared	0.038	0.101	0.263	0.374
N	3,734	3,734	3,734	3,734

Table 15: Social Learning in Bubbles

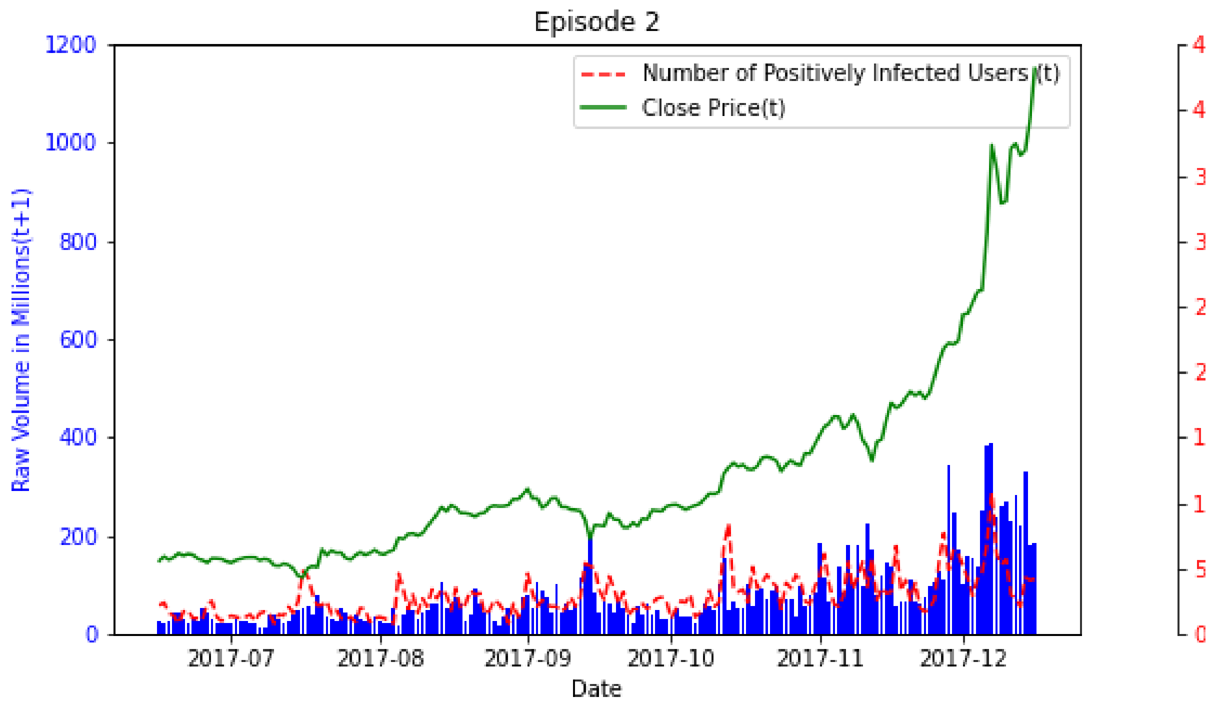
This table highlights the features of social learning in bubble. We identify bubble episodes following the procedure described in section 6.4.1. In Panel A, we describe the market conditions in bubbles. In Panel B, we highlight the features of social activities in bubbles. In Panel C, we describe sentiment contagions in bubbles. In all panels, we compare features between bubble and non-bubble episodes. We rely on seemingly unrelated regression (SUR) to test the joint significance of differences in all features. We report the joint F-statistic and its corresponding p-value. Our sample spans from May 1, 2012 to July 30, 2022.

	Bubble Episode		Non-Bubble Episode		Bubble Episode minus Non-Bubble Episode	
Features	Mean	Std	Mean	Std	Difference	t-statistic
Panel A: Market Variables						
Daily Return(Annualized)	4.050	17.66	0.491	15.72	3.56	5.352
Return Volatility(Within Day)	0.040	0.03	0.037	0.04	0.003	2.413
Total Dollar Volume (in millions)	94.071	148.80	47.503	77.63	46.568	11.784
RavenPack News Sentiment	0.168	0.42	0.040	0.43	0.127	5.894
Google Search (Bitcoin)	0.085	0.48	0.017	0.40	0.068	3.961
Panel B: Social Interactions						
Average Sentiment	0.305	0.11	0.272	0.12	0.033	6.761
Standard Deviation of Sentiment	0.657	0.05	0.672	0.05	-0.015	-6.635
Number of Posts	188.764	147.95	175.713	163.66	13.052	1.968
Number of Users	133.504	105.61	116.261	114.97	17.243	3.692
Total Number of Positive Posts	104.671	87.03	90.834	87.47	13.837	3.837
Fraction of Posts with Positive Sentiment	0.543	0.09	0.516	0.09	0.028	7.563
Fraction of Sophisticated Users	0.301	0.10	0.359	0.14	-0.058	-10.523
Panel C: Sentiment Contagion						
Number of Positively Infected Users	16.704	16.24	15.335	17.33	1.369	1.938
Number of Negatively Infected Users	3.203	4.63	3.813	5.84	-0.61	-2.629
Joint F-statistic						15.49
p-value (Probability>F)						0.00

Table 16: Social Learning and Market Crashes

This table presents the predictive power of social learning on future market declines and crashes in the Bitcoin market. The main explanatory variable is the cumulative Net SCI indicator, which measures the difference in the numbers of positively affected users and negatively affected users within the past 14 days. The positively (negatively) affected users are those who adjust their sentiment upward (downward) to align with the social sentiment in conversations. We normalize it to have a zero mean and unit standard deviation. In columns (1) to (3), the dependent variables are dummy variables for returns that equal one if the cumulative return within a given horizon falls below zero. In columns (4) to (6), the dependent variable is a crash indicator, taking a value of 1 if, within a given horizon, at least one daily return falls below the 5th percentile of daily returns. Control variables are identical to those in Table 14. We report the average marginal effects. We report t-statistics in parentheses based on standard errors clustered by day. The coefficients are multiplied by 100.

	Dummy (Return < 0)			Crash (below 5% perc.)		
	(1)	(2)	(3)	(4)	(5)	(6)
	1 day	7 days	14 days	1 day	7 days	14 days
Cumulative Net SCI	1.813*	4.539***	6.228***	1.038***	5.696***	9.157***
in Past 14 Days	(1.88)	(4.74)	(6.52)	(2.87)	(7.43)	(10.05)
Bitcoin Volatility	-3.349	-23.316***	-14.375*	12.148***	45.829***	78.608***
in Past 14 Days	(-0.41)	(-2.91)	(-1.82)	(4.40)	(6.02)	(8.65)
RavenPack News Sentiment	-0.663	2.167	-5.177***	0.016	-2.286	1.848
in Past 14 Days	(-0.35)	(1.13)	(-2.73)	(0.02)	(-1.41)	(1.02)
Forum Average Sentiment	-5.163	-3.890	2.261	-0.098	5.083*	8.525**
in Past 14 Days	(-1.41)	(-1.07)	(0.63)	(-0.06)	(1.68)	(2.56)
Bitcoin Return	-9.877**	-21.082***	-9.197**	-0.522	4.478	14.380***
in Past 14 Days	(-2.34)	(-4.77)	(-2.17)	(-0.32)	(1.26)	(3.56)
Adjusted R-Squared	0.002	0.009	0.012	0.040	0.055	0.083
N	3,734	3,734	3,734	3,734	3,734	3,734



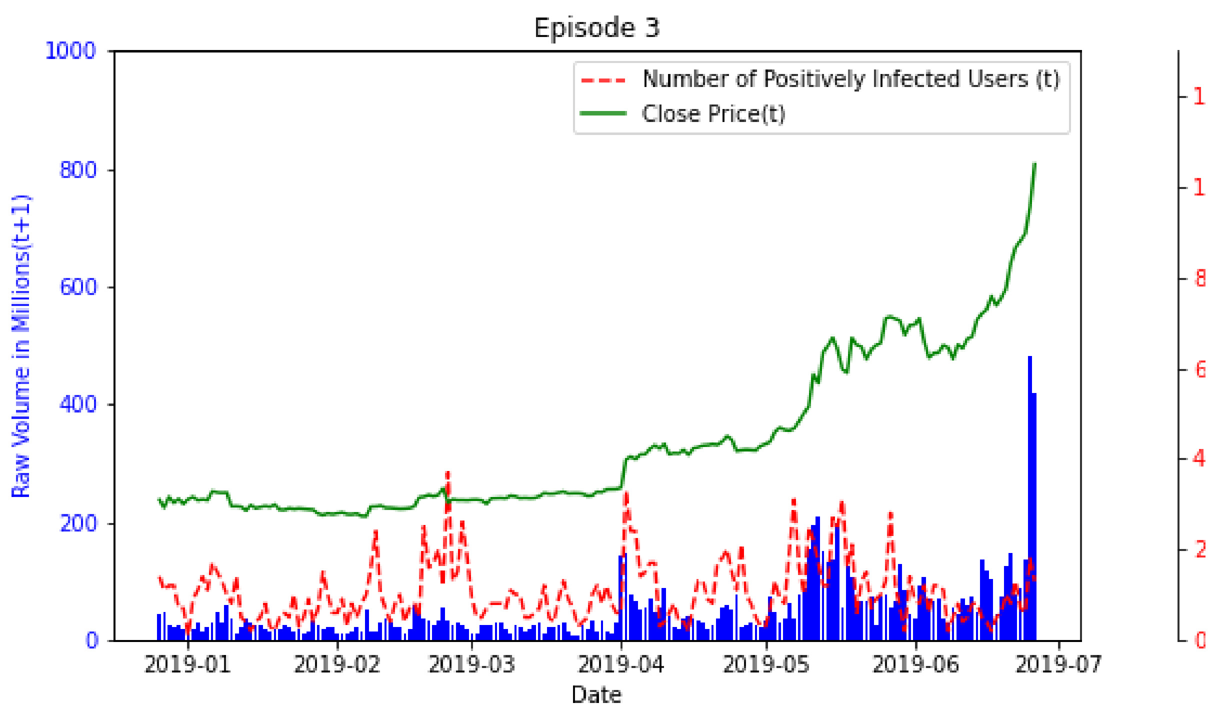


Figure A1: Sentiment Contagion and Trading Volume in Bubble Episodes

This figure plots the number of positively infected users on day t (the red dash line) and trading volume (the blue bars; in millions) on day $t+1$, along with the Bitcoin price index on day t (the green solid line). We present the second (upper panel) and thrid (lower panel) identified bubble episodes in our sample.

Table A1: Contagion Effect: Ex-Post Sentiment as Dependent Variable

This table presents the panel regression analysis of ex-post sentiment on social sentiment. The dependent variable, ex-post sentiment, refers to the sentiment in the second post of two consecutive posts (post[0] and post[1]). The main explanatory variable is social sentiment, which is the average sentiment of others' posts within the conversation published between the timestamps of post[0] and post[1]. Control variables are identical to those in Table 2. We also control for the prior sentiment. We report t-statistics in parentheses based on standard errors clustered by investor and day. We multiply each coefficient by 100.

	Ex-Post Sentiment				
	(1)	(2)	(3)	(4)	(5)
Social Sentiment	3.478***	3.474***	4.013***	3.949***	3.940***
	(9.03)	(9.03)	(8.28)	(8.20)	(8.18)
Prior Sentiment	-0.498**	-0.498**	-0.495**	-0.541**	-0.544**
	(-2.20)	(-2.20)	(-1.97)	(-2.15)	(-2.16)
RavenPack News Sentiment		0.668	0.296	0.311	0.511
between Post[0] and Post[1]		(0.94)	(0.39)	(0.41)	(0.66)
Bitcoin Return			29.835***	30.421***	31.329***
			(6.54)	(6.63)	(6.66)
Bitcoin Volatility			-1.623	-4.302	-4.900
			(-0.10)	(-0.26)	(-0.30)
Forum Sentiment				-5.862***	-5.883***
				(-6.58)	(-6.61)
RavenPack News Sentiment					1.554*
24 hours before Post[0]					(1.71)
RavenPack News Sentiment					-0.954
48 hours before Post[0]					(-1.05)
User FE	YES	YES	YES	YES	YES
Date FE	YES	YES	YES	YES	YES
Adjusted R-Squared	0.038	0.038	0.038	0.039	0.039
N	212,623	212,623	178,614	175,750	175,750

Table A2: Contagion Effect: Max of Sentiment

In this table, we use the max of sentence sentiment to measure sentiment at the post level. We present the panel regression analysis of sentiment change on social sentiment. The dependent variable is the sentiment change defined as the revision in sentiment between a user's two consecutive posts post[0] and post[1]. The main explanatory variable is the social sentiment defined as the average sentiment of others' posts within the conversation, published between the timestamps of post[0] and post[1]. Control variables are identical to those in Table 2. In all columns, we include a prior fixed effect. We report t-statistics in parentheses based on standard errors clustered by user and day. We multiply each coefficient by 100.

	Sentiment Change				
	(1)	(2)	(3)	(4)	(5)
Social Sentiment	2.562***	2.557***	2.814***	2.732***	2.726***
	(6.34)	(6.33)	(5.61)	(5.51)	(5.50)
RavenPack News Sentiment		1.111	0.404	0.400	0.463
between Post[0] and Post[1]		(1.27)	(0.43)	(0.42)	(0.49)
Bitcoin Return			36.173***	36.908***	37.514***
			(6.75)	(6.84)	(6.93)
Bitcoin Volatility			13.072	10.055	9.714
			(0.76)	(0.58)	(0.57)
Forum Sentiment				-5.306***	-5.321***
				(-5.35)	(-5.37)
RavenPack News Sentiment					1.179
24 hours before Post[0]					(1.08)
RavenPack News Sentiment					-1.758
48 hours before Post[0]					(-1.61)
Prior FE	YES	YES	YES	YES	YES
User FE	YES	YES	YES	NO	NO
Date FE	YES	YES	YES	NO	NO
Adjusted R-Squared	0.385	0.385	0.386	0.386	0.386
N	212,623	212,623	178,614	175,750	175,750

Table A3: Contagion Effect: Sum of Sentiment

In this table, we use the summation of sentence sentiment to measure sentiment at the post level. We present the panel regression analysis of sentiment change on social sentiment. The dependent variable is the sentiment change defined as the revision in sentiment between a user's two consecutive posts post[0] and post[1]. The main explanatory variable is the social sentiment defined as the average sentiment of others' posts within the conversation, published between the timestamps of post[0] and post[1]. Control variables are identical to those in Table 2. In all columns, we include a prior fixed effect. We report t-statistics in parentheses based on standard errors clustered by user and day. We multiply each coefficient by 100.

	Sentiment Change				
	(1)	(2)	(3)	(4)	(5)
Social Sentiment	2.550***	2.542***	2.849***	2.792***	2.788***
	(5.42)	(5.41)	(4.99)	(4.94)	(4.93)
RavenPack News Sentiment		1.804	1.041	1.124	0.927
between Post[0] and Post[1]		(1.33)	(0.71)	(0.76)	(0.62)
Bitcoin Return			46.987***	47.556***	47.551***
			(4.42)	(4.44)	(4.35)
Bitcoin Volatility			22.847	16.944	17.104
			(1.08)	(0.79)	(0.80)
Forum Sentiment				-5.182***	-5.194***
				(-4.73)	(-4.75)
RavenPack News Sentiment					0.379
24 hours before Post[0]					(0.23)
RavenPack News Sentiment					-3.092**
48 hours before Post[0]					(-2.03)
Prior FE	YES	YES	YES	YES	YES
User FE	YES	YES	YES	NO	NO
Date FE	YES	YES	YES	NO	NO
Adjusted R-Squared	0.310	0.310	0.311	0.311	0.311
N	212,623	212,623	178,614	175,750	175,750

Table A4: Contagion Effect: Same Conversation

This table presents the panel regression analysis of sentiment change on social sentiment when post[0] and post[1] are published in the same thread within a 24-hours window. The dependent variable is the sentiment change defined as the revision in sentiment between a user's two consecutive posts post[0] and post[1]. The main explanatory variable is the social sentiment defined as the average sentiment of others' posts within the conversation, published between the timestamps of post[0] and post[1]. Control variables are identical to those in Table 2. In all columns, we include a prior fixed effect. We report t-statistics in parentheses based on standard errors clustered by user and day. We multiply each coefficient by 100.

	Sentiment Change				
	(1)	(2)	(3)	(4)	(5)
Social Sentiment	2.165***	2.165***	1.819*	1.819*	1.821*
	(3.14)	(3.14)	(1.93)	(1.91)	(1.91)
RavenPack News Sentiment		-1.736	-1.887	-1.815	-1.657
between Post[0] and Post[1]		(-0.93)	(-0.92)	(-0.88)	(-0.81)
Bitcoin Return			33.132***	33.386***	33.676***
			(3.88)	(3.84)	(3.81)
Bitcoin Volatility			20.181	21.245	20.998
			(0.71)	(0.75)	(0.74)
Forum Sentiment				-2.369	-2.373
				(-1.52)	(-1.52)
RavenPack News Sentiment					0.662
24 hours before Post[0]					(0.33)
RavenPack News Sentiment					0.184
48 hours before Post[0]					(0.09)
Prior FE	YES	YES	YES	YES	YES
User FE	YES	YES	YES	NO	NO
Date FE	YES	YES	YES	NO	NO
Adjusted R-Squared	0.335	0.335	0.335	0.336	0.335
N	70,388	70,388	54,714	53,283	53,283

Table A5: Contagion Effect: Sentiment Change over the Next 12, 48, and 72 Hours

This table presents panel regression analysis of sentiment change on social sentiment when post[0] and post[1] are published within a 12-, 48- and 72-hours window. The dependent variable is the sentiment change defined as the revision in sentiment between a user's two consecutive posts post[0] and post[1]. The main explanatory variable is the social sentiment defined as the average sentiment of others' posts within the conversation, published between the timestamps of post[0] and post[1]. Control variables are identical to those in Table 2. In all columns, we include a prior fixed effect. We report t-statistics in parentheses based on standard errors clustered by user and day. We multiply each coefficient by 100.

	12 Hours	48 Hours	72 Hours
	(1)	(2)	(3)
Social Sentiment	2.575***	3.975***	4.235***
	(4.31)	(8.45)	(9.17)
RavenPack News Sentiment	0.627	0.220	0.192
between Post[0] and Post[1]	(0.56)	(0.29)	(0.28)
Bitcoin Return	28.611***	24.719***	21.591***
	(4.56)	(5.32)	(5.37)
Bitcoin Volatility	5.691	14.377	15.457
	(0.31)	(0.97)	(1.06)
Forum Sentiment	-3.348***	-5.744***	-6.421***
	(-2.99)	(-5.52)	(-6.17)
RavenPack News Sentiment	1.211	0.183	0.475
24 hours before Post[0]	(0.90)	(0.26)	(0.80)
RavenPack News Sentiment	-1.692	-0.806	-0.200
48 hours before Post[0]	(-1.11)	(-1.13)	(-0.33)
Prior FE	YES	YES	YES
User FE	YES	YES	YES
Date FE	YES	YES	YES
Adjusted R-Squared	0.345	0.348	0.350
N	108,372	245,944	283,588

Table A6: Contagion Effect: More Controls for Lagged Response

This table presents panel regression analysis of sentiment change on social sentiment. For column (1), we adopt the regression settings and variable definitions from column (5) in Table 2. In column (2), we include more controls to account for lagged response to past market dynamics. In column (3), we further account for lagged response to previous posts. In all columns, we include a prior fixed effect. We report t-statistics in parentheses based on standard errors clustered by user and day. We multiply each coefficient by 100.

	(1)	(2)	(3)
Social Sentiment	3.088***	3.074***	3.279***
	(6.01)	(6.02)	(6.25)
RavenPack News Sentiment between Post[0] and Post[1]	-0.032	-0.294	-0.368
	(-0.04)	(-0.33)	(-0.41)
Bitcoin Return	30.723***	34.931***	33.826***
	(5.90)	(5.67)	(5.20)
Bitcoin Volatility	12.925	17.581	9.058
	(0.81)	(1.17)	(0.56)
Forum Sentiment	-4.679***	-4.726***	-4.453***
	(-4.42)	(-4.54)	(-4.24)
RavenPack News Sentiment 24 hours before Post[0]	1.385	1.230	1.445
	(1.45)	(1.23)	(1.43)
RavenPack News Sentiment 48 hours before Post[0]	-1.030	-0.895	-0.965
	(-1.03)	(-0.87)	(-0.92)
Bitcoin Return 24 hours before Post[0]		9.400	8.376
		(1.50)	(1.30)
Bitcoin Volatility 24 hours before Post[0]		13.303	2.721
		(0.28)	(0.06)
Bitcoin Return 48 hours before Post[0]		1.016	0.092
		(0.19)	(0.02)
Bitcoin Volatility 48 hours before Post[0]		-4.064	0.960
		(-0.08)	(0.02)
Average Sentiment of Posts before Post[0]			-4.236***
			(-4.85)
Prior FE	YES	YES	YES
User FE	NO	NO	NO
Date FE	NO	NO	NO
Adjusted R-Squared	0.345	0.345	0.346
N	175,750	175,749	170,313

Table A7: Contagion Effect: Sentiment Change for Users with and without Wallet Address

This table presents a subsample analysis of sentiment change on social sentiment. Specifically, we divide the user sample into two groups: users who do not disclose their wallet address and users who voluntarily reveal their wallet address. For each subsample analysis in columns (1) to (2), we adopt the regression settings and variable definitions from column (5) in Table 2. We investigate the interactive effect in column (3). In all columns, we include a prior fixed effect. The t-statistics (in parentheses) are based on standard errors clustered by user and day. We multiply each coefficient by 100.

	Users Without Wallet	Users With Wallet	Interactive Effect
	(1)	(2)	(3)
Social Sentiment	3.088***	3.611***	3.057***
	(5.31)	(2.72)	(5.31)
User with Wallet * Social Sentiment			0.187
			(0.14)
RavenPack News Sentiment	-0.608	2.538	-0.032
between Post[0] and Post[1]	(-0.62)	(1.14)	(-0.04)
Bitcoin Return	30.580***	30.938**	30.722***
	(5.73)	(2.08)	(5.90)
Bitcoin Volatility	20.176	-54.355	12.920
	(1.23)	(-1.08)	(0.81)
Forum Sentiment	-4.457***	-5.377**	-4.678***
	(-3.84)	(-2.03)	(-4.42)
RavenPack News Sentiment	0.569	5.236**	1.385
24 hours before Post[0]	(0.56)	(2.27)	(1.45)
RavenPack News Sentiment	-2.111*	3.058	-1.030
48 hours before Post[0]	(-1.86)	(1.26)	(-1.03)
Prior FE	YES	YES	YES
User FE	YES	YES	YES
Date FE	YES	YES	YES
Adjusted R-Squared	0.344	0.341	0.345
N	146,974	28,405	175,750

Table A8: Social Learning and Individual Trading: Between Consecutive Pairs

This table presents the regression analysis of individuals' trading decisions on the sentiment change between post[0] and post[1]. The dependent variable is an indicator of whether the individual actively increased her portfolio by at least 0.1 Bitcoins. The main explanatory variable is the sentiment change, defined as the revision in sentiment between a user's two consecutive posts post[0] and post[1]. Control variables are identical to that in Table 2. Additionally, we include the time gap between t_0 and t_1 as an additional control, as trading is more likely to occur over longer time windows. In all columns, we include user and date fixed effects, as well as a prior fixed effect. We report t-statistics in parentheses based on standard errors clustered by user and day. We multiply each coefficient by 100.

	Probability buy in [t0, t1]	Probability buy in [t0, t1],t1-t0>2 hours	Probability buy in in [t1, t1+7days]
Sentiment Change	8.035***	10.229***	0.436*
	(2.71)	(3.35)	(1.96)
Prior FE	YES	YES	YES
User FE	YES	YES	YES
Date FE	YES	YES	YES
Controls	YES	YES	YES
Adjusted R-Squared	0.238	0.250	0.632
N	203	189	9,284

Table A9: Post Decisions and Recent Bitcoin Returns

This table presents the association between users' post decisions and Bitcoin returns in the past 14 days. In columns (1) and (2), the dependent variable $PostDecisions_{i,t}$ is a dummy variable that equals one if user i makes at least one post on day t and zero otherwise. We focus on users' active days after their registration. In columns (3) to (6), we investigate how, conditional on users posting, recent Bitcoin returns affect the post sentiment. In columns (3) and (4), the dependent variable $PositivePosts_{i,t}$ equals to one if user i publishes a positive posts on day t and zero otherwise. In columns (5) and (6), the dependent variable $PessimisticPosts_{i,t}$ equals to one if user i publishes a negative posts on day t and zero otherwise. We calculate the average levels of sentiment in RavenPack news in the past 14 days as controls for news arrivals. We control for market information by controlling for two Bitcoin market variables: Bitcoin Volatility and Number of Transactions. Each of two variable is the cumulative sum in the past 14 days. We calculate forum average sentiment as the average sentiment of posts on the Bitcointalk forum published in the past 14 days. In all columns, we add user fixed effect. We multiply each coefficient by 100. We report t-statistics in parentheses based on standard errors clustered by user and date.

	(1)	(2)	(3)	(4)	(5)	(6)
	Post	Post	Positive	Positive	Pessimistic	Pessimistic
	Decisions	Decisions	Posts	Posts	Posts	Posts
Bitcoin Return	0.210***	0.299***	6.676***	2.549***	-4.920***	-2.373***
in Past 14 Days	(3.96)	(4.92)	(10.00)	(3.81)	(-9.54)	(-4.59)
Bitcoin Volatility		1.016***		2.728**		0.047
in Past 14 Days		(11.12)		(2.28)		(0.05)
RavenPack News Sentiment		-0.042***		2.446***		-1.986***
in Past 14 Days		(-3.15)		(10.33)		(-11.86)
Forum Sentiment		-0.231**		43.072***		-26.911***
in Past 14 Days		(-2.08)		(19.20)		(-15.59)
User FE	YES	YES	YES	YES	YES	YES
Date FE	NO	NO	NO	NO	NO	NO
Controls	NO	YES	NO	YES	NO	YES
Adjusted R-Squared	0.055	0.055	0.039	0.039	0.020	0.023
N	122,046,111	121,971,272	649,710	403,816	649,710	403,816

Table A10: Fraction of Positive and Negative Posts and Recent Bitcoin Returns

This table presents the association between the fraction of positive and negative posts in the forum within a day and Bitcoin returns in the past 14 days. In columns (1) and (2), the dependent variable is the fraction of positive posts on the forum within a given day. In columns (3) and (4), the dependent variable is the fraction of negative posts in the forum within one day. We calculate the average levels of sentiment in RavenPack news in the past 14 days as controls for news arrivals. We control for Bitcoin Volatility which is based on the returns in the past 14 days. We multiply each coefficient by 100. We report t-statistics in parentheses based on standard errors clustered by date.

	(1)	(2)	(3)	(4)
Bitcoin Return in Past 14 Days	5.434***	4.382***	-4.545***	-3.803***
	(6.03)	(5.41)	(-7.73)	(-7.28)
Bitcoin Volatility in Past 14 Days		-9.334***		6.638***
		(-8.17)		(9.03)
RavenPack News Sentiment in Past 14 Days		2.448***		-1.731***
		(7.56)		(-8.12)
Adjusted R-Squared	0.014	0.042	0.025	0.060
N	3,743	3,743	3,743	3,743

Table A11: Optimism in Conversations and Future Market Crashes

In this table, we study how optimism in conversations predicts the occurrence of crashes using the probit model. The dependent variable is a future crash indicator, which takes on a value of 1 if there is any return crash in the next K hours (K =6, 24, 48, and 72) and 0 otherwise. We classify the event of return crash in a given hour if the future Bitcoin return within the subsequent 24 hours is less than -15%. Our key predictor for the future crash is the Optimism variable, and we define it in two distinct ways. In columns (1), (3) and (5), Optimism in summation is the aggregated sentiment in the current hour, where the aggregated sentiment is simply the summation of sentiment across posts published within an hour. In columns (2), (4) and (6), Optimism in fraction is the fraction of hours with positive aggregated sentiment in the past 7*24 hours, where the aggregated sentiment is simply the summation of sentiment across posts published within an hour. We report the average marginal effects. We report t-statistics in parentheses based on standard errors clustered by day. Finally, we multiply each coefficient by 100.

	(1)	(2)	(3)	(4)	(5)	(6)
	6 Hours	6 Hours	24 Hours	24 Hours	72 Hours	72 Hours
Optimism in Summation	0.179*		0.419**		1.312***	
	(1.95)		(2.16)		(3.58)	
Optimism in Fraction		0.414**		1.182***		2.782***
		(2.24)		(3.24)		(4.38)
RavenPack News Sentiment	0.039	-0.184	0.519	-0.215	1.002	-0.751
in Past 7*24 Hours	(0.06)	(-0.29)	(0.44)	(-0.19)	(0.50)	(-0.39)
Bitcoin Return Volatility	170.494***	167.886***	441.053***	422.925***	1021.164***	981.951***
in Past 7*24 Hours	(4.78)	(4.76)	(6.16)	(6.27)	(8.16)	(8.18)
Cumulative Bitcoin Return	-4.347	-3.308	-12.444*	-9.780	-18.378	-10.586
in Past 7*24 Hours	(-1.02)	(-0.78)	(-1.65)	(-1.27)	(-1.44)	(-0.82)
Adjusted R-Squared	0.109	0.116	0.127	0.141	0.140	0.153
N	17,953	17,953	17,953	17,953	17,953	17,953

Table A12: Optimism in Conversations and Future Volatility

In this table, we study how optimism in conversations predicts future number of transactions volume. The dependent variable volatility is the standard deviation of hourly returns in the next K hours (K = 6, 24, 72). We adjust the standard deviation to the daily level. Our key predictor for the future volatility is the Optimism variable, and we define it in two distinct ways. In columns (1), (3) and (5), Optimism in summation is the aggregated sentiment in the current hour, where the aggregated sentiment is simply the summation of sentiment across posts published within an hour. In columns (2), (4) and (6), Optimism in fraction is the fraction of hours with positive aggregated sentiment in the past 7*24 hours, where the aggregated sentiment is simply the summation of sentiment across posts published within an hour. We multiply each coefficient by 100. In this Table, we use Bootstrap standard errors from a simulation of 1000 times.

	(1) 6 Hours	(2) 6 Hours	(3) 24 Hours	(4) 24 Hours	(5) 72 Hours	(6) 72 Hours
Optimism in Summation	0.288***		0.260***		0.221***	
	(9.29)		(10.37)		(10.83)	
Optimism in Fraction		0.207***		0.215***		0.203***
		(6.75)		(9.01)		(10.44)
RavenPack News Sentiment	-0.351***	-0.442***	-0.121	-0.232**	0.137	0.023
in Past 7*24 Hours	(-2.96)	(-3.52)	(-1.12)	(-2.02)	(1.55)	(0.25)
Bitcoin Return Volatility	426.657***	433.016***	405.186***	409.712***	363.787***	366.767***
in Past 7*24 Hours	(28.08)	(28.43)	(29.94)	(30.06)	(32.86)	(32.71)
Cumulative Bitcoin Return	-17.934**	-16.988**	-15.784***	-14.915***	-8.160**	-7.411**
in Past 7*24 Hours	(-2.57)	(-2.43)	(-3.26)	(-3.07)	(-2.39)	(-2.17)
Adjusted R-Squared	0.286	0.282	0.328	0.324	0.331	0.328
N	17,953	17,953	17,953	17,953	17,953	17,953

Table A13: Correlations between SCI and Other Indicators

This table shows the correlations between SCI indicator and other variables of interest measured at daily frequency. SCI indicator measures the daily number of users who adjust their sentiment to align with the social sentiment in conversations. We use its logarithmic form and eliminate time-fixed effects by regressing on year-month indicators and weekday dummies. Forum Average Sentiment reflects the average level of sentiment in posts published on the forum. RavenPack News Sentiment represents the daily mean sentiment level in news documented in the RavenPack news database. RavenPack News Coverage denotes the number of news articles regarding Bitcoin within a single day. Google Search Bitcoin records the detrended Google Search Index for Bitcoin on a daily basis. Our sample spans from May 1, 2012 to July 30, 2022.

Variables	SCI	Forum Average Sentiment	RavenPack News Sentiment	RavenPack News Coverage	Google Search Bitcoin
SCI	1.000				
Forum Average Sentiment	0.006	1.000			
	(0.696)				
RavenPack News Sentiment	-0.003	0.347	1.000		
	(0.876)	(0.000)			
RavenPack News Coverage	0.004	-0.085	-0.082	1.000	
	(0.842)	(0.000)	(0.000)		
Google Search (Bitcoin)	0.238	-0.032	0.024	0.296	1.000
	(0.000)	(0.050)	(0.265)	(0.000)	