Social Interactions and Lottery Stock Mania*

Turan G. Bali
David Hirshleifer
Lin Peng
Yi Tang
Qiguang Wang

This draft: August 2025

Abstract

We find that social interactions are associated with stocks becoming more lottery-like and with greater investor overoptimism about the lottery characteristic. Heightened social media activity about a stock positively predicts the probability of an extreme daily price run-up, a *lottery event*. Lottery event stocks subject to more extensive social media discussions subsequently experience greater retail buying pressure—particularly from Robinhood users—followed by lower returns. Moreover, lottery stocks of firms headquartered in more socially connected counties experience lower subsequent returns. Our findings are consistent with theories in which social interactions stimulate investor excitement and asset price bubbles.

JEL classification: G10, G11, G12, G14, C13, E20, E30.

Keywords: speculation, lottery stocks, meme stocks, social interactions, social media, investor attention, retail investors, MAX

*This paper draws upon parts of a previous working paper that was circulated under the title "Attention, Social Interaction, and Investor Attraction to Lottery Stocks." Turan G. Bali: McDonough School of Business, Georgetown University, email: turan.bali@georgetown.edu. David Hirshleifer: Marshall School of Business, University of Southern California, email: hirshlei@marshall.usc.edu. Lin Peng: Zicklin School of Business, Baruch College/CUNY (corresponding author, lin.peng@baruch.cuny.edu). Yi Tang: Gabelli School of Business, Fordham University, email: ytang@fordham.edu. Qiguang Wang: Hong Kong Baptist University, Hong Kong, China, email: qiguangw@hkbu.edu.hk. We thank Hank Bessembinder, Huaizhi Chen (discussant), Ozgur Demirtas, Xi Dong, Joseph Engelberg (discussant), Marcel Fafchamps, Jon Garfinkel, Sanjev Goyal, Danling Jiang, Lei Jiang, Patrick Konermann, Dan Kovenock, Jens Kvarner, Weikai Li (discussant), Eric Lie, Yueran Ma, Yiming Qian, Larry Samuelson, Tyler Shumway (discussant), Johannes Stroebel, Adam Szeidl, Paul Tetlock, Alex Wolitzky, Kamil Yilmaz, Jianfeng Yu, Xiaoyan Zhang, Dexin Zhou, and Hao Zhou and seminar and conference participants at the American Finance Association Meeting, the Miami Behavioral Finance Conference, Baruch College, BI Norwegian Business School, Koc University, Rutgers University, Sabanci University, the European Networks Conference at the University of Essex, NYUAD Networks, Markets, State and Conflict conference, the State University of New York, Stony Brook, the Tsinghua Finance Workshop, the University of Iowa, and the University of International Business and Economics for helpful comments. Lin Peng acknowledges the Krell research grant for financial support and the Chinese University of Hong Kong ShenZhen for the data support she received while she was a visiting scholar.

1. Introduction

There is extensive evidence that investors—especially retail investors—are attracted to stocks with lottery-like payoffs, resulting in high valuations of such assets and low subsequent returns.[1] Recent years have seen the rise of a modern incarnation of lottery investing in the form of meme stocks such as GameStop, AMC, and Palantir. Speculation in lottery-like assets, including options, meme tokens, and cryptocurrencies, has intensified during 2024 and early 2025.[2] Meme stocks are also associated with the gamification and democratization of finance, as exemplified by Robinhood, a zero commission retail trading platform.

In remarks that seem to reflect these trends, Warren Buffet's annual shareholder letter released on February 24, 2024 drew parallels between today's stock market and a casino (Buffett 2024): "For whatever reasons, markets now exhibit far more casino-like behavior than they did when I was young." Buffet further noted, "The casino now resides in many homes and daily tempts the occupants."

The rise of meme stocks has gone hand in hand with activity on popular investor social media platforms such as StockTwits and Reddit.[3] Stories about retail investors becoming millionaires overnight from investing in meme stocks and other speculative assets have sparked extensive discussions on social media platforms. While previous studies on lottery stocks have primarily focused on the role of fixed and inherent preferences for lottery characteristics at the individual investor level (Barberis and Huang 2008; Brunnermeier and Parker 2005), these recent developments raise the questions of how social interactions

[1] See, e.g., Kumar (2009), Boyer, Mitton, and Vorkink (2010), Bali, Cakici, and Whitelaw (2011), Green and Hwang (2012), Barberis, Mukherjee, and Wang (2016), and Liu et al. (2020).

[2] GameStop, BlackBerry and online pet-products retailer Chewy reportedly increased more than 90% in the 12 months leading up to February 2025. Trump and Melania Trump meme coins reached market capitalizations of approximately $15 billion and $2 billion, respectively, since their January 2025 debuts (Krystal Hur, "Investors Spot Signs of market Forth During Long Bull Market," *Wall Street Journal*, February 16, 2025). Stocks popularly referred to as meme stocks have repeatedly appeared in the top decile when ranked by the magnitude of their largest one-day returns in a month (a measure of lotteriness, Figure 1).

[3] Several recent studies investigate the role of social media in finance (see, e.g., Antweiler and Frank 2004, Chen et al. 2014, Giannini, Irvine, and Shu 2019, Cookson and Niessner 2020, Hu et al. 2021, Chen and Hwang 2022, Farrell et al. 2022, Cookson, Engelberg, and Mullins 2023).

direct investor attention and enthusiasm toward speculative investments (Hirshleifer 2020, Han, Hirshleifer, and Walden 2022; Pedersen 2022).

This paper provides insights into these questions using data from two types of social networks: a leading investment social media platform, StockTwits, and a general-interest social media platform that is representative of the general population, Facebook. Using data from the investment social media platform, we find that the amount of social media discussion about a stock predicts both the likelihood of an extreme daily price run-up (a *lottery event*) and its magnitude. Among stocks that exhibit prominent lottery features (high "lotteriness"), higher social media activity predicts more aggressive retail buying and greater overvaluation, as evidenced by substantially lower subsequent returns and more negative price reactions to future earnings announcements. These stocks also experience greater investor disagreement and higher trading volume.

Furthermore, the finding for the general interest social media platform reinforces the insights derived from the investment social media platform, showing that lottery stocks of firms headquartered in counties with stronger social connectivity experience greater overpricing. Together, these results based on both types of social media platforms suggest that social mechanisms play a crucial role in the emergence, pricing dynamics, and trading patterns of lottery stocks.

Our tests are motivated by theoretical models proposing that social interactions can propagate incorrect beliefs and naïve allocation of investor attention.[4] Burnside, Eichenbaum, and Rebelo (2016) provide an epidemiological model in which home buyer beliefs with differing optimism spread via social interactions, resulting in housing booms and busts. In the model of Hirshleifer (2020), stock market bubbles are caused by the social transmission of folk mental models about market fundamentals. As applied to mental

[4]Several models have shown that social interactions can spread rumors and amplify behavioral biases (DeMarzo, Vayanos, and Zwiebel 2003, Hirshleifer 2020, Han, Hirshleifer, and Walden 2022), trigger information cascades (Bikhchandani, Hirshleifer, and Welch 1992, Banerjee 1992), create free-riding incentives (Han and Yang 2013), and introduce "mutations" and transmission failures (Jackson, Malladi, and McAdams 2024). In addition, theories and evidence suggest that when investors exhibit Keeping-Up-with-the-Joneses preferences, social interactions can induce herding into risky securities and trigger excessive trading of local stocks (DeMarzo, Kaniel, and Kremer 2008; Gomez, Priestley, and Zapatero 2009; Roussanov 2010; Hong et al. 2014; Bali et al. 2023).

models about the prospects of firms for big successes, this implies price bubbles in lottery stocks. In Pedersen (2022), as investment ideas propagate through social interactions, influential "fanatic" agents substantially affect equilibrium price owing to their influence on listeners. The interaction between rational agents and fanatics within the social interaction over time leads to attraction to speculative stocks, price bubbles, and subsequent reversals. Such effects are reminiscent of that observed for meme stocks, such as the GameStop bubble in 2021.

The bubble dynamics in these models indicate that greater social interaction can promote extreme (though temporary) price run-ups. If we measure lotteriness by the occurrence of an extreme run-up, then such models imply that social interactions can cause a stock to become a lottery stock.

Similarly, in the social transmission model of Han, Hirshleifer, and Walden (2022), social interactions lead to positive investor feedback in response to high past returns, thereby intensifying upside price movements. Furthermore, in this model, stocks with high skewness are overpriced and on average earn low subsequent abnormal returns. This occurs because high returns are disproportionately reported by signal senders and highly salient for signal receivers, leading the latter to become overoptimistic about the stocks they hear about.

Alternative approaches to understanding mispricing of lottery stocks are based on non-standard preferences for lottery-like payoffs (Brunnermeier and Parker 2005; Barberis and Huang 2008) or on beliefs that overweight the probability of extreme, salient payoff outcomes (Bordalo, Gennaioli, and Shleifer 2012, 2013). These theories assume that investors know the relevant payoff distributions (i.e., the lottery features) of stocks. In reality, investors have limited attention and do not continually monitor all stocks for lotteriness, which varies over time. Investors form beliefs about payoff distribution either by directly observing stock characteristics and outcomes, or by communicating with others. It follows that social interactions can amplify preference- or salience-based effects.

By directing investor attention to these features, these interactions can therefore increase investor demand and induce greater overpricing of such stocks.

Based on these considerations, we hypothesize that social interactions contribute to the pricing and trading of lottery stocks in two ways. First, for a stock with high lotteriness, social interactions can stimulate optimism about the stock's future payoffs, and can direct investor attention to the stock, both of which imply retail buying and overpricing. Second, social interactions can contribute to a stock becoming more lottery-like. Once this occurs, the first channel can then promote buying. So via both channels, we predict that investor attraction to and market overpricing of lottery stocks increases with the extent of social interactions among potential investors.

To test this hypothesis, we construct two key measures: a measure of the lotteriness of a stock and a measure of the extent of social interaction. Our primary measure of a stock's lotteriness follows the approach of Bali, Cakici, and Whitelaw (2011). We identify days on which a stock's return on that day is the highest over a trailing 21-trading-day window. We denote such days as MAX events and their corresponding returns as MAXRET. The magnitude of MAXRET captures a stock's lotteriness. We also verify robustness using the alternative lottery index proposed by Kumar (2009).

Our primary proxy for the extent of social interactions—word-of-mouth discussions in investor social networks—derives from activity on StockTwits, the largest finance-specific social media platform where investors share their opinions.[5] We apply a dataset comprising over 10.7 million messages posted by 937,564 users between 2010 and 2022, covering 12,325 distinct stocks. We measure social media engagement by the daily volume of StockTwits messages referencing a specific stock.

We use these measures to investigate the hypothesis that social interactions promote high valuations of lottery stocks. We first document a sharp increase in message activity referencing a stock around lottery events (defined as days with extreme one-day returns,

[5]This dataset has been increasingly used in recent studies, including Cookson and Niessner (2020), Cookson et al. (2024), and Cookson, Engelberg, and Mullins (2023), and is reviewed in Cookson, Mullins, and Niessner (2024).

i.e., MAXRET in the top 10% cross sectionally). This suggests the possibilities that these discussions triggered such events, that such events are highly salient and heavily discussed by investors in their online social interactions, or both.

We then assess whether lottery stocks are overpriced by testing whether lotteriness negatively predicts returns, and whether social interactions influence the sensitivity of subsequent returns to a stock's lotteriness. Previous studies have shown that extreme positive one-day returns on a stock negatively predict future returns in the following month. To test for the effects of social interactions, we compare one-month-ahead-returns for stocks with MAX events, focusing on those with elevated StockTwits message volume in the days leading up to the event versus those without such activity.

We find that a one-standard-deviation increase in MAXRET combined with elevated message volume, is associated with much lower one-month-ahead returns—a difference of 136 basis points. Notably, we find that message volume, when interacted with MAXRET, fully absorbs the predictive power of MAXRET itself in forecasting future returns. We also show that the moderating role of social interactions in the lottery characteristics–return relation is distinct from traditional attention measures capturing retail and institutional activity. This strongly suggests that social interactions are a key contributor to the overvaluation of lottery stocks, as implied by the social-based models of Hirshleifer (2020), Han, Hirshleifer, and Walden (2022), and Pedersen (2022).

Another standard approach to testing for and estimating overpricing is to test for price corrections at subsequent earnings announcement dates. If lottery stocks are overpriced, the release of fundamental news should trigger corrections that are commensurate with the initial level of overpricing. We find that lottery stocks on average experience negative abnormal returns in the days around the next quarter's earnings announcement dates, with more negative returns for stocks that had higher message volume prior to the MAX day. These findings provide further support for the hypothesis that more extensive social interactions promote more over-optimistic expectations about lottery stocks.

We then probe underlying mechanisms byexploring how social media discussions influence the likelihood of a stock experiencing a lottery event and whether these discussions affect investors' subsequent trading behavior. We first show that social media discussion directly predicts the extent to which a stock becomes more lottery-like. Stocks that become the focus of extensive discussions on StockTwits are 65% more likely to experience a lottery event the following day relative to the mean. Furthermore, among stocks that experience a lottery event, those with extensive pre-event discussions have an average event-day return that is 17% higher than the unconditional mean. We exploit the staggered adoption of state-level social media privacy laws as an exogenous shock to social interactions and show that our results are causal. This finding suggests that social media contributes to a stock's likelihood of achieving lottery status by directing investor attention to such stocks and sparking their optimism about the stock's future prospects.

Retail investors, in particular, are more likely to be drawn to the salient features of lottery stocks and are more susceptible to the influence of social media. We therefore expect that social interactions will especially bolster demand for lottery stocks among retail investors. We test this hypothesis using two types of data: trading activity on the Robinhood platform and aggregate retail order flows.

The Robinhood platform has been the go-to platform for retail investors central to the meme stock phenomenon. Using stock popularity data from the Robintrack website, we investigate whether StockTwits message volume around a lottery event predicts Robinhood buy herding episodes, defined as a day with a sharp increase in the number of Robinhood users holding a stock. We find that lottery stocks with more extensive discussion on StockTwits are 38% more likely to experience a Robinhood buy herding episode in the week following the lottery event, relative to the unconditional mean.

To assess whether our findings about StockTwits messages and buy herding on Robinhood are representative of retail traders more broadly, we examine aggregate retail order imbalances using measures proposed by Boehmer et al. (2021) and Barber et al. (2024). We find that, conditional on a lottery event, heavier discussion on StockTwits is associated

with greater aggregate retail net purchases in the following week. This evidence suggests that our findings generalize to the broader retail investor population, where greater social media discussion is linked to more aggressive retail buying of lottery stocks.

Theories of social interactions further suggest that such interactions can spread rumors, propagate incorrect beliefs, and introduce signal mutation, thereby creating scope for investor disagreement and trading.[6] We therefore test whether greater StockTwits message activity around lottery events is associated with greater investor disagreement and trading volume. These dimensions of investor beliefs and trading have not been explored in previous empirical studies of lottery stocks.

We find that lottery stocks with elevated message volume on the days leading up to the lottery event experience more message sentiment disagreement over the following two weeks—by 6.4% and 7.9% of the sample mean—relative to those without such message volume. These stocks also see substantially higher share turnover during the same period, by 67% and 70%, relative to the sample mean, respectively. These findings provide further evidence that social interactions influence belief updating following lottery events.

Although StockTwits is the largest finance-specific social media platform, it only constitutes a subset of investors, raising the question of how representative StockTwits is of the thinking and behavior of investors at large. Our finding that StockTwits activity predicts real trading outcomes suggests that the information extracted from the platform is not a mere sideshow. To further address generalizability, we use the large-scale Facebook-based social network data to construct an alternative proxy for the extent of social interaction among investors.

The Facebook's Social Connectedness Index (SCI, Bailey et al. 2018b) measures friendship likelihood between Facebook users across counties. As the world's largest online social network, Facebook's enormous scale and coverage (over 258 million monthly active users in the US as of 2020) and the relative representativeness of its user base make SCI a unique measure of the real-world geographic structure of US social networks at the popu-

[6]See, for example, Shiller (2000), Hirshleifer (2020), Pedersen (2022), Jackson, Malladi, and McAdams (2024), and Hirshleifer, Peng, and Wang (2024).

lation scale (Bailey et al. 2018a,b; Chetty et al. 2022). Compared to the investing-focused StockTwits-based measure, SCI better captures long-run, real-world social ties between counties. The two measures thus complement each other in testing our hypotheses.

Using the Facebook data, we investigate how the social connectedness of a county where a lottery stock's headquarters is located influences investor demand for the stock. Extensive evidence shows that investors are more attentive to nearby firms and are more likely to invest in and trade these firms' stocks (see, e.g., Coval and Moskowitz 1999, Huberman 2001). As investors discuss their gains from investing in local stocks, particularly those with lottery-like characteristics, news about these stocks spreads from local investors to those in other counties through word-of-mouth. Consequently, such stocks tend to be extensively discussed among a broader range of investors. We therefore expect that lottery stocks based in counties with higher social connectivity will attract greater investor interest, leading to increased overvaluation.

To test this, we define a firm's headquarters social connectedness (SCIH) as the average connectedness of the firm's headquarters to all other U.S. counties. Higher SCIH implies that a lottery stock is more likely to be discussed by investors across the U.S., triggering stronger demand for the stock. Consistent with this, we find that lottery stocks headquartered in high-SCIH counties experience more negative returns over the next three months, suggesting that greater social connectivity amplifies the overvaluation of these stocks. The consistency of our findings across both the large-scale Facebook-based network data and the more granular StockTwits-based investor network reinforces this conclusion.

Our evidence of social interactions promoting investor attraction to lottery stocks is primarily concentrated among small stocks. This is consistent with retail investors being more prone to lottery-related biases and being especially important for trading in small, volatile stocks. However, retail investors have become increasingly important in financial markets overall, representing a rapidly growing share of market trading activity.[7] Fur-

[7]By 2021, retail trading in the U.S. accounts for almost as much volume as mutual funds and hedge funds combined (Martin and Wigglesworth 2021). At the start of 2023, U.S. retail investor daily inflows reached a record-breaking high of $1.5 billion—more than double the pre-2019 figure of just over $600 million (Rao 2023).

thermore, even large-cap stocks, such as Amazon, Facebook, Google, Nvidia, and Tesla, have also qualified as lottery stocks at various points in our sample, drawing substantial retail investor attention.[8] Previous research shows that lottery demand is important for several important stylized facts about asset pricing, such as the betting-against-beta anomaly (Bali et al. 2017), the asset growth anomaly (Lu et al. 2022), and for the preference of certain actively managed U.S. equity funds to hold lottery stocks (Agarwal, Jiang, and Wen 2021). These findings underscore the importance of understanding the drivers behind the rise of lottery stocks and investors' demand for such stocks.

Our paper contributes to the aforementioned literature on the overpricing of lottery characteristics and associated negative return predictability by providing key pieces of evidence consistent with the newer, social-based theories about the formation of lottery stock bubbles and investor behavior during such episodes. Our findings also provide insights into the more general literature on retail investors as a possible source of stock market anomalies (Barber and Odean 2008, Da, Engelberg, and Gao 2011, Kelley and Tetlock 2013, 2017, Andrei and Hasler 2015, Yuan 2015, Atilgan et al. 2020, and Boehmer et al. 2021), including research on Robinhood investors (Welch 2022, Barber et al. 2022, Ozik, Sadka, and Shen 2021 and Eaton et al. 2022). Furthermore, Green and Jame (2024) find that such retail trading frenzies are predictive of equity issuance and increased real investment. Our evidence suggests that retail attention is socially transmitted, and that social networks act as a conduit that prompts the herding and purchasing of speculative stocks.

2. Data, Variable Definitions, and Preliminary Tests

Our sample consists of US common stocks (SHRCD = 10 or 11) traded on the NYSE, AMEX, and NASDAQ for the period from June 1963 through December 2022. We obtain

[8] In the online appendix S.I, we provide a list of stocks in the top decile of NYSE market capitalization that obtain lottery status. For the 2010−2022 sample period, there are 148 stock-month of these large-cap lottery stock observations, with the mean and median average daily abnormal Google search volume (ASV) over the $[-11, 0]$ pre-lottery event window at 20% and 17%, respectively, indicating a substantial increase in search activity relative to the base volume. In addition, 122 stock-month observations exhibit abnormal StockTwits activity, with their pre-event message counts ranking in the top 10% cross-sectionally.

investor social media data from StockTwits for 2010–2022, stock data from the Center for Research in Security Prices (CRSP), and other accounting and financial statement variables from the merged CRSP-Compustat database.

2.1. The Lottery Characteristic and Returns to Lottery-Based Portfolios

Our measure of a stock's lotteriness follows the approach of Bali, Cakici, and Whitelaw (2011) (BCW), who use a stock's maximum daily return in a given month to measure its lottery-like characteristics for that month and predict the following month returns.[9] However, since we study daily variations in lotteriness in relation to social media and trading activities, the original BCW definition cannot be directly applied for return prediction. For example, for tests at a daily frequency, it would be inappropriate on January 15 to use the highest return during all of January to predict returns for January 16, as this procedure would introduce look-ahead bias.

We therefore introduce a variation of BCW's definition that is suitable for daily tests. Specifically, we define a MAX event day as when a stock's return is the highest over a trailing 21-trading-day window, with MAXRET representing the return on that day. A lottery event occurs if MAXRET falls in the top decile of the daily cross-sectional distribution. This approach ensures that our return predictor relies only on past data.

To assess whether this modified maximum daily return measure, MAXRET, exhibits a return predictive ability similar to that of the original BCW measure, we perform an analysis parallel to that of Bali, Cakici, and Whitelaw (2011) for the period spanning June 1963 to December 2022, replacing the original BCW measure with MAXRET. Specifically, for each month, we sort stocks into decile portfolios based on each stock's highest

[9] An alternative way to define a stock's lotteriness uses three characteristics: low price, high idiosyncratic volatility, and high estimated idiosyncratic skewness (volatility and skewness estimated over the preceding 6 months; Kumar 2009). Since these characteristics have limited daily variation, the maximum daily return-based lotteriness measure is better suited for our settings. Previous studies that study lottery-like securities include Kumar, Page, and Spalt (2011, 2016), Green and Hwang (2012), Annaerta, DeCeustera, and Verstegena (2013), Han and Kumar (2013), Boyer and Vorkink (2014), Walkshausl (2014), Barberis, Mukherjee, and Wang (2016), Zhong and Gray (2016), Carpenter, Lu, and Whitelaw (2021), Bali et al. (2017), Wang, Yan, and Yu (2017), An et al. (2020), Lin and Liu (2018), and Liu et al. (2020).

MAXRET value for that month and compute the value-weighted portfolio returns for the subsequent month. Table A1, Panel A shows that the return spreads for the long-short portfolios are negative and highly significant, ranging from 82 to 120 basis points per month. These spreads are driven primarily by the underperformance of high-MAXRET stocks. For comparison, untabulated results show that the long-short portfolio based on the original BCW measure generates return spreads between 73 to 117 basis points, which are also driven by the underperformance of stocks with high BCW values. Panels B and C parallel the Bali, Cakici, and Whitelaw (2011) Fama-MacBeth regression results using the MAXRET and the BCW measures, respectively, and provide similar results.[10]

Additionally, Online Appendix Table S.II reproduces Appendix Table A1, Panel A, for the sample of firms with above-NYSE-median market capitalization. It demonstrates that the negative return predictability of MAXRET remains robust even for the relatively large stocks.

Overall, our analysis paralleling the BCW tests confirms that the MAXRET measure of a stock's lottery characteristics has monthly predictive power that is similar to the original BCW measure. Henceforth, we rely on the MAXRET measure to capture daily variations in lotteriness and to examine its relationship with social media activity, trading behavior, and stock returns.

2.2. StockTwits Data

Our primary measure for the extent of social interaction uses data from StockTwits, a leading social media platform where investors share their investment experiences and opinions. Similar to Twitter, StockTwits allows users to post short messages, initially limited to 140 characters until May 8, 2019, when the limit was expanded to 1,000 characters. What sets StockTwits apart from Twitter is its exclusive focus on financial markets. The platform was the first to introduce the cashtag notation, using the ticker symbol of a stock in a message (e.g., $TSLA for Tesla).

[10]We conduct robustness checks using the lottery index adapted from Kumar (2009), which has a correlation of approximately 0.6 with MAXRET, and report the results in Subsection 6.3. Our findings are also robust to alternative definitions of lottery events, as detailed in Subsection 6.4.

Using the StockTwits API, we compile a dataset of over 10.7 million messages posted by 937,564 users between 2010 and 2022, covering 12,325 distinct stocks.[11] We define *Message* as the number of messages mentioning a stock during a given time period, measuring the extent of investor social interactions regarding that stock. Higher message counts indicate greater social engagement among investors.

2.3. Other Variables

We obtain the daily and monthly return and volume data from CRSP. We adjust stock returns for delisting to avoid survivorship bias (Shumway 1997). Unless otherwise stated, all variables are measured as of the end of the previous month of a MAX event so that there is no look-ahead bias in our empirical tests.

We use a number of well-known cross-sectional return predictors as control variables. Specifically, following Fama and French (1992), we estimate stock i's market beta (BETA), and compute the stock's size (SIZE) as the product of the price per share and the number of shares outstanding (in millions of dollars). The book-to-market equity ratio (BM) at the end of June of year t is computed as the book value of stockholders' equity, plus deferred taxes and investment tax credit (if available), minus the book value of the preferred stock for the last fiscal year ending in $t-1$, scaled by the market value of equity at the end of December of $t-1$. Following Fama and French (2015) and Hou, Xue, and Zhang (2015), a stock's investment related characteristic is measured by the annual growth rate of total assets (IA) at the end of June of year t is measured by the change in book assets (Compustat item AT) for the last fiscal year ending in year $t-1$ divided by lagged AT. Similarly, quarterly operating profitability (OP) is determined by income before extraordinary items (item IBQ) for the most recent fiscal quarter, with quarterly earnings announcements made in or prior to, but no longer than six months before the portfolio formation month, divided by one-quarter-lagged book equity. Momentum (MOM) is the cumulative return of a stock over a period of 11 months ending one month prior to the portfolio formation month (Jegadeesh and Titman 1993). Following Jegadeesh (1990), we

[11] The API is available at https://api.StockTwits.com/developers.

include past one-month return (REV) to capture the short-term reversal effect. We define a stock's monthly co-skewness (COSKEW) following Harvey and Siddique (2000). Following Amihud (2002), a stock's monthly illiquidity (ILLIQ, scaled by 10^6) is the average daily ratio of the absolute stock return to the dollar trading volume.[12]

2.4. Preliminary Tests and Summary Statistics

We first provide preliminary evidence that StockTwits message activity is strongly associated with a stock's lotteriness. Our tests are performed at the individual stock-day level, centered around each stock's MAX event days. Each day, we sort stocks that experienced MAX events into decile portfolios based on their MAXRET, with Decile 10 corresponding to the events with the top 10% MAXRET in the cross-section (i.e., lottery stocks).

Figure 2 displays the average daily number of StockTwits messages for each decile portfolio around the MAX events (day 0). Notably, there is a sharp increase in message counts on day 0 for lottery stocks (in Decile 10), rising from below 50 messages prior to the event day to nearly 150 messages on the MAX day. These elevated message activity persists for up to five days following the event. This suggests that the lottery events are highly salient and heavily discussed online. The message activity for the lottery stocks rises even before the MAX event, particularly on the day immediately preceding the event. In comparison, stocks in Decile 9 experience only a slight increase in message activity and there are no significant changes in message activity around the MAX events for stocks in the other deciles.

[12]We require a minimum of 24 monthly observations for market beta and co-skewness, computed from monthly data over the past five years, and a minimum of 15 daily observations for Amihud illiquidity, computed from daily data in a month. Depending on availability, we use the redemption, liquidation, or par value (in that order) to estimate the book value of the preferred stock. Following Bali, Brown, and Tang (2017), our data for the operating profitability variable (OP) begins in 1972. This reflects the limited availability of earnings announcement dates (Compustat item "RDQ") prior to 1972, a critical variable necessary for properly lagging OP relative to future stock returns and avoiding potential look-ahead bias. Following Gao and Ritter (2010), we adjust for institutional features of the way that the NASDAQ and NYSE/AMEX volumes are counted. Specifically, we divide the NASDAQ volume by 2.0, 1.8, 1.6, and 1 for the periods prior to February 2001, between February 2001 and December 2001, between January 2002 and December 2003, and January 2004 and later years, respectively.

We next present summary statistics for the key variables of interest and the controls. Panel A of Table 1 provides the time-series averages of the cross-sectional descriptive statistics for the aforementioned variables. It shows that the mean (median) MAXRET is 6.42 (4.41) percentage points, and there is substantial cross-sectional variation, with a standard deviation of 9.32 percentage points. The average MAXRET 90^{th}-percentile breakpoint (above which an event is defined as a lottery event) is 11.96 percentage points.

Regarding StockTwits message activity, we examine the variable Messages, which corresponds to the number of messages posted from day -11 up to, and including, the MAX event day (denoted as day 0). The distribution of the variable is highly right-skewed, with a median value of 12, mean of 108.17, and an average skewness of 214.0. There tend to be very low message counts for most of the MAX event observations, whereas for popular stocks in the top 10th percentile, the average message counts exceeds 123. Given this, we define an indicator variable, $\mathbb{1}_{mssg}$, which takes the value of one if message counts for the $[-11, 0]$ window rank above the 90th percentile of its cross-sectional distribution, and zero otherwise. Stocks with $\mathbb{1}_{mssg}$ equal to one are presumed to be subjected to significantly more extensive social media discussions compared to those with $\mathbb{1}_{mssg}$ equal to zero.[13]

Panel B of Table 1 presents the time-series averages of the cross-sectional correlations among the variables used in this study. Focusing on $\mathbb{1}_{mssg}$, it shows that the variable is associated with lower one-month-ahead returns, with an average cross-sectional correlation of -1.9%. Additionally, $\mathbb{1}_{mssg}$ is positively correlated with stock characteristics such as MAXRET, market beta, size, investment growth, momentum, and co-skewness, and negatively correlated with book-to-market.

In our subsequent tests, we investigate how measures of social interaction are associated with the returns and trading of lottery stocks, while controlling for stock characteristics that earlier studies have shown to predict future returns and trading activities.

[13] In Section 6.4, we demonstrate that our findings remain robust when using alternative windows to define $\mathbb{1}_{mssg}$.

3. Social Interactions and Returns to Lottery-like Stocks

As discussed in the introduction, social transmission bias can cause investors to be attracted to lottery stocks (Han, Hirshleifer, and Walden 2022) and communication can generate the "GameStop"-like price bubbles (Pedersen 2022). Furthermore, social interactions can amplify the visibility of a stock's lottery-like characteristic, thereby magnifying the effect on investor demand of inherent preferences for lottery features or of investor attention to salient payoffs (Barberis and Huang 2008; Bordalo, Gennaioli, and Shleifer 2012, 2013; Brunnermeier and Parker 2005). In this section, we test the pricing implications of the proposed mechanisms and examine whether social interactions amplify the overvaluation of lottery stocks by examining post-event returns and price reactions to subsequent earnings announcements.[14]

3.1. Lottery Stock Returns

Our hypothesis is that social interactions contribute to the high valuations of stocks with high levels of lotteriness. We therefore predict that such stocks will experience lower mean abnormal returns in the following months as prices correct. We test this hypothesis by adopting an event study framework using panel regressions. This approach takes advantage of daily variations in a stock's lotterienss and the dynamic nature of StockTwits activities.

We first assess whether the panel regression methodology generates results comparable to those obtained with the Fama-MacBeth approach used by Bali, Cakici, and Whitelaw (2011). As a baseline specification, we estimate the following panel regressions of returns following a stock's MAX event:

$$\mathrm{R}_{i,t+1} \quad = \quad \lambda_0 + \lambda_1 \mathrm{MAXRET}_{it} + \lambda_2 \mathrm{X}_{it} + \varepsilon_{i,t+1}, \tag{1}$$

[14] More broadly, a substantial literature documents peer effects on investment decisions (see Kelly and O'Grada (2000), Duflo and Saez (2002, 2003), Hong, Kubik, and Stein (2004, 2005), Brown et al. (2008), Cohen, Frazzini, and Malloy (2008), Shive (2010), Kaustia and Knüpfer (2012), Pool, Stoffman, and Yonker (2015), Heimer (2016), Ahern (2017), Crawford, Gray, and Kern (2017), Maturana and Nickerson (2018), Hong and Xu (2019), Ouimet and Tate (2020), and Huang, Hwang, and Lou (2021)). Social interactions have also been found to promote the sale of lottery tickets (Mitton, Vorkink, and Wright 2018). Kuchler and Stroebel (2021) provide a review of this literature.

where R is the cumulative returns for stock i over 21, 42, and 63 trading days post-MAX event. MAXRET is the stock return on the MAX event day. $\mathbf{X}$ is a vector of lagged control variables, following Bali, Cakici, and Whitelaw (2011), Fama and French (2015) and Hou, Xue, and Zhang (2015).[15] We include firm and year-month fixed effects to account for omitted firm characteristics and marketwide shocks and cluster the standard errors by firm and year-month.

The regression is estimated over the period 2010–2022, for which the StockTwits data are available. The results, presented in Table A2 in the Appendix, show that the coefficient of MAXRET is negative and highly significant, consistent with the Fama-MacBeth results (Panel B of the Appendix Table A1). Given the advantages of the panel regression methodology—its applicability to daily event-study setting and its ability to incorporate firm and time fixed effects to account for omitted factors—our subsequent tests employ panel regressions unless otherwise noted.

As discussed in the introduction, we hypothesize that social interactions contribute to the pricing of lottery stocks by: (1) stimulating optimism about a stock with high lotteriness, and (2) making a stock more lottery-like. To test the first of these hypotheses, we next examine whether StockTwits message activity in days leading up to a MAX event influences the relationship between MAXRET and the stock's subsequent returns. This test requires measurement of both message activity prior to the MAX event date and on the MAX event date, which we define using the $[-11, -1]$ window and day 0, respectively. Accordingly, we define a high-message indicator, $\mathbb{1}_{mssg}$, which equals one if the number of StockTwits messages during the $[-11, 0]$ window belongs to the top 10 percent of the distribution, and zero otherwise.

Extending equation (1), we estimate the following panel regression:

$$\mathrm{R}_{i,t+1} = \lambda_0 + \beta \mathrm{MAXRET}_{it} \cdot \mathbb{1}_{mssg_{it}} + \lambda_1 \mathrm{MAXRET}_{it} + \lambda_2 \mathbb{1}_{mssg_{it}} + \lambda_3 \mathrm{X}_{it} + \varepsilon_{i,t+1}. \quad (2)$$

[15] $\mathbf{X}$ includes the market beta (BETA), the natural logarithm of firm size (SIZE), the natural logarithm of the book-to-market ratio (BM), total asset growth rate (IA), operating profitability (OP), momentum (MOM), past one-month return (REV), illiquidity (ILLIQ), and co-skewness (COSKEW), as detailed in Section 2.3.

The variable of interest is the interaction term, MAXRET$\cdot\mathbb{1}_{mssg}$, which captures how StockTwits message activity modulates the relation between MAXRET and subsequent returns. As before, we include firm and year-month fixed effects to account for omitted firm characteristics and cluster the standard errors by firm and year-month.

The results are presented in Table 2. Column 1 shows that the coefficient for the interaction term, MAXRET$\cdot\mathbb{1}_{mssg}$, is -0.146 (t-statistic $= -3.55$) for the [1, 21] post-MAX return. This negative relationship persists for longer horizons, with coefficients ranging from -0.170 to -0.248 for 42- and 63-day cumulative returns (columns 2 and 3).[16] Economically, a one standard-deviation increase in MAXRET is associated with an additional 136 to 231 basis point decrease in future returns over one to three months for stocks with high StockTwits activity.

A key finding is that including $\mathbb{1}_{mssg}$ and its interaction with MAXRET fully absorbs the effect of MAXRET itself—the MAXRET coefficients are negative but substantially smaller in absolute magnitude and are no longer significant. These findings, coupled with the highly significant negative slope coefficients for MAXRET$\cdot\mathbb{1}_{mssg}$, suggest that social interactions could be a key factor contributing to the underperformance of lottery stocks in our sample.

3.2. Subsequent Earnings Announcement Returns

Another way to test for overvaluation is to estimate stock price corrections at the dates of subsequent earnings announcements. If lottery stocks are overpriced, the arrival of fundamental news should induce price corrections with magnitudes commensurate with initial overvaluation levels. We therefore hypothesize that these corrections will be more pronounced for lottery stocks that experienced higher levels of social interaction activity.

[16]Control variables SIZE, IA, OP, REV, and COSKEW show significant coefficients consistent with prior literature (Fama and French, 2015; Hou, Xue, and Zhang, 2015; Jegadeesh, 1990; Harvey and Siddique, 2000). Robustness tests using the database of Cookson and Niessner (2023), available 2010–2021, yield similar results, with MAXRET$\cdot\mathbb{1}_{mssg}$ coefficients of -0.122, -0.181, and -0.257 for one-, two-, and three-month-ahead returns (Table A3). We thank Tony Cookson and Marina Niessner for sharing their data.

To test this prediction, we calculate the DGTW (1997) cumulative abnormal returns over the three days surrounding the earnings announcement, denoted as CAR (see, for example, Frazzini 2006; Kaniel et al. 2012). We then estimate equation (2) with the next fiscal quarter CAR as the dependent variable. The results are presented in Table A4.

As a benchmark for comparison, we first present the regression without the StockTwits message indicator in column 1. The coefficient of MAXRET is -3.010 and highly significant, indicating that high MAXRET stocks tend to experience disappointing earnings announcements in the following quarter. These results are consistent with Engelberg, McLean, and Pontiff (2018), who find increased returns to anomaly portfolios on earnings announcement days and other corporate news days. They argue that this phenomenon indicates that information arrival plays a role in rectifying investors' biased expectations.

We then introduce the StockTwits message indicator and its interaction term with MAXRET, with message activities measured over the $[-11, 0]$ window. The variable of interest is the interaction term, MAXRET$\cdot\mathbb{1}_{mssg}$. In column 2, the coefficient is significantly negative, with a value of -2.517. This indicates that MAX events associated with very high pre-event message volume tend to be followed by earnings announcement returns in the next quarter that are 122% ($= -2.517/-2.069$) more negative than the rest of the MAX events.

This result supports our hypothesis that extensive social media discussions in the days leading up to a MAX event promote optimistic investor beliefs about the lottery stock, resulting in greater disappointment upon the future release of earnings news. This evidence is consistent with the implications of the social interaction models of Hirshleifer (2020), Han, Hirshleifer, and Walden (2022) and Pedersen (2022).

A caveat to our analysis is that the message activity indicator is defined over the [−11, 0] window, as motivated by the theoretical models discussed earlier. The problem is that omitted variables could be correlated with both MAX event date message activity and MAX event date returns, potentially driving our results. The inclusion of firm and day fixed effects, as well as a rich list of known return predictors, mitigate the problem by ex-

cluding the possibility that time-invariant firm characteristics or marketwide factors drive our results. To go further, we extend the analysis by considering a plausibly exogenous shock to investor social interactions, applying staggered implementation of state-level social media privacy laws as an exogenous shock to social media activity in Section 6.1. To provide further insight and to further address potential endogeneity concerns, in the next section we also explore the underlying mechanisms for our main findings.

4. Mechanisms

In this section, we explore the implication of theoretical models that social interactions can cause a stock to become a lottery stock and that social interactions trigger more aggressive retail buying of lottery stocks. We therefore test whether social interactions predict the likelihood of lottery event occurrence and investor trading behavior around these events. Since these are predictive tests, they are not subject to endogeneity deriving from correlated contemporaneous variables nor from reverse causality from the dependent variable to the independent variable.

4.1. The Emergence of Lottery Stocks

We begin by testing whether StockTwits message activity predicts both the probability of stocks attaining lottery status and the extremity of their lottery characteristics.

Previous studies of lottery stocks have mostly taken the stock's lotteriness as given and have focused on subsequent returns. There has been little exploration of the predictors of becoming a lottery stock and why lottery characteristics change over time.[17] Motivated by the models of Hirshleifer (2020), Han, Hirshleifer, and Walden (2022) and Pedersen (2022), we propose that social interactions contribute to a stock becoming more lottery-like, and that social interactions that direct investor attention to such stocks stimulate overoptimism about its future payoffs. Our evidence in Figure 2, showing a substantial rise in message activity even before a large MAXRET event, is consistent with this perspective.

[17]Previous studies have examined the extent to which investors' preference for lottery attributes is influenced by persistent socioeconomic factors such as education, religion, and household income (Kumar 2009; Han and Kumar 2013; Bali et al. 2023), and preference for lottery-like payoffs (Boyer, Mitton, and Vorkink 2010; Bali, Cakici, and Whitelaw 2011).

We formally test this hypothesis by examining the extent to which social interactions, as measured by StockTwits message counts, contribute to both the likelihood of a stock becoming lottery-like (the extensive margin) and the magnitude of its lotteriness (the intensive margin).

To investigate the extensive margin, we estimate the following panel regressions with all stock-day observations in our sample:

$$\mathbb{1}_{i,t+1}^{lottery} = \lambda_0 + \lambda_1 \mathbb{1}_{mssg,it}^{preEvent} + \lambda_2 \mathrm{X}_{it} + \varepsilon_{i,t+1}, \tag{3}$$

where $\mathbb{1}^{lottery}$, the lottery event indicator, takes a value of one if a stock experiences a lottery event for a given day (day 0), and zero otherwise. The primary independent variable of interest is the pre-MAX event message indicator, $\mathbb{1}_{mssg}^{preEvent}$, which equals one if the message counts during the $[-11, -1]$ window exceeds the top 10% of the cross-sectional distribution, and zero otherwise. $\mathbf{X}$ denotes Return, the cumulative return during the $[-11, -1]$ window, and the vector of control variables specified in equation (2), lagged by one month. We include firm and year-month fixed effects and use standard errors clustered by firm and year-month.

The results are presented in Table 3. The slope coefficient of $\mathbb{1}_{mssg}^{preEvent}$ is 0.228 in the univariate panel regression (column 1), and 0.272 after controlling for all other variables (column 2). Both coefficients are highly significant. The economic magnitude can be illustrated based on the coefficient estimate in column 2: an increase in message activity to the top decile increases the likelihood of a stock obtaining the lottery status on a given day by 0.272 percentage points, a substantial magnitude compared to the unconditional probability of 0.42 percentage points. The results indicate that elevated levels of social interaction activity is a strong predictor of the emergence of lottery stocks.[18]

[18]The coefficient of R^{preEvent} is significantly negative for the full sample of all stock-days (columns 1−2) but positive for the lottery event sample (columns 3−4). This is because, for the full sample, high returns during $[-11, -1]$ reduce the likelihood of even higher returns on day 0, making lottery events less probable. In contrast, lottery event days by construction have returns exceeding those in the $[-11, -1]$ window, explaining the positive R^{preEvent} coefficients in columns 3−4. Our findings remain robust when using an alternative, path-independent lottery event definition, as shown in Subsection 6.4.

We next investigate the intensive margin, that is, the association between StockTwits message activity and the magnitude of a stock's MAXRET on lottery event days. We estimate the following stock-month panel regressions linking the magnitude of MAXRET with the pre-lottery event message indicator ($\mathbb{1}_{mssg}^{preEvent}$):

$$\text{MAXRET}_{i,t+1} = \lambda_0 + \lambda_1 \mathbb{1}_{mssg,it}^{preEvent} + \lambda_2 \text{X}_{it} + \varepsilon_{i,t+1}, \quad (4)$$

where $\mathbf{X}$ denotes the vector of control variables specified in equation (3). We also include firm and year-month fixed effects.

Columns 3 and 4 of Table 3 present the results, with two-way clustered standard errors by firm and year-month. The coefficient on $\mathbb{1}_{mssg}^{preEvent}$ is positive and highly significant, suggesting that more extensive social media discussion is associated with higher lottery-event day returns. Specifically, the slope coefficient on $\mathbb{1}_{mssg}^{preEvent}$ in column 4 implies that a stock that has experienced elevated StockTwits message flows before the lottery event is expected to have substantially higher levels of MAXRET, by 3.867 percentage points, which is 17% higher than the cross-sectional lottery-day mean return of 23.0 percentage points.

In summary, these results show that StockTwits message activity strongly predicts both the magnitude of large one-day share price run-ups and the likelihood that a stock becomes a lottery stock. This evidence is consistent with the predictions of the theories of Hirshleifer (2020), Han, Hirshleifer, and Walden (2022) and Pedersen (2022) that social interactions can increase lotteriness and induce overvaluation of lottery stocks. In the next two subsections, we complement the return tests by examining investors' trading activity around lottery events.

4.2. Retail Herding on the Robinhood Trading Platform

We next test the hypothesis that social interactions trigger more aggressive retail buying of lottery stocks. Specifically, we examine whether StockTwits message activity

around the time of lottery events predicts greater retail investor buying, particularly by Robinhood investors.

A notable modern type of lottery stock is the "meme stock," whose rise in early 2021 is closely tied to investors on the Robinhood trading platform. Robinhood pioneered commission-free trading, making stock market participation more accessible to a wider audience, especially younger and first-time investors. This feature was crucial in attracting a large user base who were eager to trade without the financial barriers traditionally associated with brokerage services. The platform has also introduced several means of gamifying stock trading. An example is that "new members were given a free share of stock, but only after they scratched off images that looked like a lottery ticket."[19]

Consequently, Robinhood emerged as the preferred trading platform for retail investors who played a pivotal role in the meme stock phenomenon. The number of active users on the platform grew from 0.5 million in 2014 to 12.5 million by 2020. Barber et al. (2022) examine stocks on Robinhood's Top Movers list, comprising of stocks with the largest absolute percentage price changes from the previous-day close. They find that Robinhood investors engage in attention-induced buying more than other retail investors, leading to increased correlation in their purchase behavior and contributing to the overvaluation of the corresponding stocks.

Motivated by these findings and the model of Han, Hirshleifer, and Walden (2022), which suggests that attraction to a stock can be socially transmitted, we investigate whether social interactions promote the purchase of lottery stocks by Robinhood investors.

We use stock popularity data from the Robintrack website for the period between May 2, 2018 and August 13, 2020. Following Barber et al. (2022), we define a herding episode as a day when the daily increase in the number of Robinhood users for a stock ranks within the top 5% of the cross-sectional distribution of the lottery stock sample, requiring a minimum threshold of 100 new users on that day. This represents approximately 14 herding episodes per day, or 0.5% of the average number of stocks covered by the Robin-

[19]Nathaniel Popper, "Robinhood Has Lured Young Traders, Sometimes with Devastating Results," *The New York Times*, July 8, 2020.

track data on a given day, comparable to the 0.5% threshold used by Barber et al. (2022) for the full Robintrack sample.[20]

We first present graphical evidence of a stock's herding propensity around lottery events. We classify stock-day observations into two groups: the High-message group, corresponding to days when message counts belong to the top 10% of the cross-sectional distribution; and the Low-message group, which includes all other days. Figure 3 displays the frequencies of herding episodes for these two groups, as defined by the prior-day message counts. The figure indicates that 43.2% of lottery events with elevated prior-day discussions on StockTwits are associated with a Robinhood herding episode, substantially exceeding the 19.1 percentage points likelihood observed for lottery events with lower message volume. Similar patterns are observed on the days surrounding the lottery event day.

We describe statistically how StockTwits message activity around a lottery event day (day 0) relates to subsequent herding episodes using the following panel regression:

$$\mathrm{HR}[n]_{it} \quad = \quad \alpha_0 + \alpha_1 \mathrm{Message}_{it} + \alpha_2 \mathrm{X}_{it} + \varepsilon_{it}, \tag{5}$$

where n corresponds to the lottery event day (day 0) or the post-lottery event windows of [1, 5] and [6, 10]. The herding indicator HR equals one if a stock experiences a herding episode on day 0 (HR[0] in column 1) and zero otherwise. For columns 2 and 3, HR$[1,5]$ and HR$[6,10]$ represent the averages of daily herding indicators over days [1, 5] and [6, 10], respectively.

Table 4 presents the results. The key variable of interest is the StockTwits message activity indicator, Message. In column 1, Message is the pre-event message volume indicator $\mathbb{1}_{mssg}^{preEvent}$, which equals one if a stock's message count over $[-11,-1]$ ranks within the top 10% cross-sectionally, and zero otherwise. In columns 2 and 3, Message is the event-day message indicator $\mathbb{1}_{mssg}^{0}$, which equals one if the stock's day-0 message count

[20]Due to data limitations, we do not examine herding episodes specifically associated with meme stocks in the post-2020 period.

ranks within the top 10% cross-sectionally, and zero otherwise. $\mathbf{X}$ includes a vector of control variables following Barber et al. (2022).[21] The regression includes year-month fixed effects, and standard errors are clustered by firm and year-month and t-statistics are shown in parentheses. We do not include firm fixed effects as an average firm has only three lottery-event observations during the sample period with available Robintrack data.

The results show that the coefficients on the high-message indicator $\mathbb{1}_{mssg}^{preEvent}$ in column 1 is 0.070 and statistically significant. This indicates that stocks with high message volumes are 7.0 percentage points more likely to experience a Robinhood herding episode—equivalent to 27% of the average herding probability of 26.3 percentage points on the lottery event days.[22]

Similarly, the significant coefficient of 0.020 for $\mathbb{1}_{mssg}^{preEvent}$ in column 2 indicates that lottery events accompanied by elevated StockTwits activity raise the next-week Robinhood buy herding probability by 2.0 percentage points—38% of the 5.3 percentage point baseline herding probability for that week. Column 3 shows a positive but insignificant association between message activity and herding probability over days [6, 10].

Overall, these findings indicate that more intense social media activity concerning lottery stocks predicts increased buy herding among Robinhood investors around lottery event days. This result suggests that social interactions play an important role in the attention-driven herding documented by Barber et al. (2022).

4.3. Aggregate Retail Trading

To assess whether the findings for Robinhood investors are indicative of retail investors more generally, we next examine the behavior of aggregate retail order flows around lottery

[21] The control variables include: return (Return), herding indicator (HR), log number of Robinhood users (USER), log change in users (ΔUSER), abnormal trading volume (AbVol), and abnormal news (AbNews). AbVol is the natural logarithm difference between daily share volume and its past 21-day mean. AbNews is the log difference between one plus daily Ravenpack news count and its past 21-day mean. For column 1, Return is the cumulative return over days $[-11, -1]$, and other controls are averaged over the same period. For columns 2 and 3, all controls are measured as of day 0. The earnings announcement indicator (EA) equals one if there is an earnings announcement on the MAX day, and zero otherwise. Given the short length of the panel, 27 months, and limited time-series variation, we omit firm fixed effects (Petersen 2009).

[22] Among other explanatory variables, column 1 shows that past return, lagged herding status, and Robinhood user numbers positively predict herding likelihood, consistent with Barber et al. (2022).

events. Following Boehmer et al. (2021) and Barber et al. (2024), we define the volume-based retail order imbalance (OIB) for stock i on day d as follows:

$$\text{OIB}_{id} = \frac{\text{BVOL}_{id} - \text{SVOL}_{id}}{\text{BVOL}_{id} + \text{SVOL}_{id}}, \tag{6}$$

where BVOL and SVOL represent the number of shares bought and sold by retail investors for stock i on day d, respectively.[23]

To test whether social interactions affect retail trading of lottery stocks, we modify equation (5) by using retail OIB as the dependent variable and by adding lagged OIB as a control. Table 5 presents the findings, with Panels A and B corresponding to the Robinhood and the StockTwits sample, respectively.

Panel A shows that the coefficients on $\mathbb{1}_{mssg}^{preEvent}$ and $\mathbb{1}_{mssg}^{0}$, the primary variables of interest, are positive and significant in predicting OIBs on day 0 and days [1, 5], respectively. Specifically, stocks subject to more extensive discussions on StockTwits on days [−11, −1] experience 2.66 percentage points higher net retail buys on day 0. Similarly, stocks subject to more extensive discussions on day 0 experience 1.47 percentage points higher net retail buys in the subsequent week. These increases represent 10.2% and 9.1% of their respective sample standard deviations. This effect is primarily observed during days [1, 5]. It is considerably weaker and is statistically insignificant over days [6, 10], as presented in column 3.

Panel B extends the analysis to the entire StockTwits universe, which includes eight times the number of observations used in Panel A. This panel omits the Robinhood-specific explanatory variables. While pre-event messages do not significantly predict event-day retail order imbalance (column 1), message activity more strongly predicts retail net

[23]The methodology uses off-exchange trades (with an exchange code equal to "D") from the TAQ database. A trade is classified as a retail buy if the sub-penny transaction price is above the quote midpoint and as a retail sell if the transaction sub-penny price is below the quote midpoint. The mean and standard deviation of OIB on the lottery event day are 0.02 and 0.26, respectively. In comparison, the corresponding mean and standard deviation of daily OIB for the sample of Robinhood stocks during the period from May 2, 2018 to August 13, 2020, as reported by Barber et al. (2022), are 0.01 and 0.35, respectively. We are grateful to Barber et al. (2024) for sharing the code used to calculate retail order imbalance.

buying in the first two weeks after the lottery event (columns 2 and 3). The delayed OIB response in the StockTwits sample, compared to the Robinhood sample, may reflect slower information diffusion about lotteriness, particularly among less visible stocks in the broader StockTwits universe.

Overall, the evidence indicates that the insights obtained from the Robinhood investor tests generalize to retail investors more broadly, including those who do not use Robinhood. Together, in both investor samples, elevated StockTwits message activity predicts more aggressive net retail buying of lottery stocks.

4.4. Disagreement and Trading Volume

We have shown that social interactions among investors are associated with aggregate returns and trading activities. In this subsection, we take advantage of the rich, message-level, information in our StockTwits data and explore unique implications of theories based upon social interaction in the context of lottery stocks.

For example, theories of social interactions suggest that word-of-mouth communication in social interactions can spread rumors, incorrect beliefs, or naïve trading strategies (Shiller 2000, Hirshleifer 2020, Han, Hirshleifer, and Walden 2022, Hirshleifer, Peng, and Wang 2024). Furthermore, Jackson, Malladi, and McAdams (2024) show that, even for rational individuals, message relaying can introduce "mutations". These effects give scope for investor disagreement regarding asset valuations. Empirically, there is evidence that social interactions trigger echo chamber effects among investors (Cookson, Engelberg, and Mullins 2023) and are linked to persistent disagreement and trading volume following earnings announcements (Hirshleifer, Peng, and Wang 2024).

Motivated by these studies, we examine how StockTwits message activity relates to investor disagreement and trading volume around lottery events. We obtain message sentiment and define disagreement (DIS) as the standard deviation of sentiment probabilities across messages for a stock-day.[24]

[24]We measure message sentiment with self-labeled bullish or bearish indicators provided by users (43.6% of messages). For messages without self-labels, we apply the Bidirectional Encoder Representations from Transformers (BERT), a state-of-the-art model that has demonstrated superior performance in sentiment

Table 6 presents the results from the following panel regression using daily observations around lottery events:

$$\text{DIS}[n]_{it} \quad = \quad \alpha_0 + \alpha_1 \text{Message}_{it} + \alpha_2 \text{X}_{it} + \varepsilon_{it}, \tag{7}$$

where n represents the lottery event day (day 0) or the post-lottery event windows of [1, 5] and [6, 10]. The dependent variable DIS, in percentage points, is message disagreement on day 0 (column 1) or average daily disagreement for days [1, 5] and [6, 10] (columns 2 and 3).

The key variable of interest is the StockTwits message activity indicator, Message. The corresponding measures, $\mathbb{1}_{mssg}^{preEvent}$ and $\mathbb{1}_{mssg}^{0}$ indicate high message count during the pre-event window $[-11, -1]$ or on day 0, respectively, as defined in Table 4. $\mathbf{X}$ represents a vector of control variables.[25] We include firm and year-month fixed effects and cluster standard errors by firm and year-month.

The coefficients on the message indicators, $\mathbb{1}_{mssg}^{preEvent}$ and $\mathbb{1}_{mssg}^{0}$, are consistently positive across all columns and highly significant. In terms of economic magnitude, Column 1 indicates that lottery stocks that attracted high StockTwits activity over days [−11, −1] experience 0.016 higher investor disagreement on day 0, or 6.6% of the sample mean. Similarly, high message activity on day 0 predicts 6.4% and 7.9% higher disagreement during days [1, 5] and [6, 10], respectively. These results suggest that social interactions intensify disagreement on the StockTwits platform.[26]

classification (Devlin et al. 2019). Using self-labeled messages as the training sample, we use BERT to calculate a continuous sentiment score that range from 0 (bearish) to 1 (bullish). The mean and stardard deviation of sentiments are 0.65 and 0.42, respectively.

[25]Following Table 4, we include the following controls: Return, AbVol, AbNews, and EA. DIS is average daily message disagreement over days [−11, −1] for column 1 and daily message disagreement on day 0 for columns 2 and 3. As in Table 2, we also control for the market beta (BETA), log market capitalization (SIZE), log book-to-market (BM), total asset growth rate (IA), operating profitability (OP), momentum (MOM), past one-month return (REV), illiquidity (ILLIQ), and co-skewness (COSKEW).

[26]For the control variables, day 0 disagreement is positively associated with abnormal volume, and abnormal news over days $[-11, -1]$, and is higher for growth stocks (low BM) and past winners (MOM). Columns 2 and 3 show that the predictive power of these variables persists for the next two weeks. Additionally, the coefficient on Return is also positive for columns 2 and 3, indicating that an extremely large return on day 0 positively predicts investor disagreement over the next two weeks.

An extensive literature proposes that investor disagreement is associated with trading volume (see, e.g., Kim and Verrecchia 1991, Harris and Raviv 1993, Kandel and Pearson 1995, and Scheinkman and Xiong 2003). We therefore examine the association of StockTwits message activity with the trading volume around lottery event days. Specifically, we estimate equation (7) with share turnover (TO, in percentage points) as the dependent variable, adding lagged TO as an additional control. Share turnover is measured as traded shares divided by shares outstanding. The mean of TO on lottery event days is 17.7 percentage points. For days [1, 5], the mean and standard deviation of daily TO are 2.1 and 2.3 percentage points, respectively.

The results are presented in Table 7. Column 1 shows that the coefficient of $\mathbb{1}_{mssg}^{preEvent}$ is 10.7 and highly significant. This indicates that lottery stocks with high StockTwits activity during days $[-11, -1]$ exhibit 60.5% higher trading volume on day 0 relative to the sample mean. Similarly, columns 2 and 3 show that high message activity on day 0 corresponds to 66.7% and 70.4% higher trading volume (relative to the mean) over the following two weeks.

Our findings indicate that greater StockTwits discussion of lottery stocks predicts elevated trading in these stocks. To the best of our knowledge, this paper is the first to document a strong relationship between message activity, investor disagreement, and trading volume in the context of lottery stocks. This evidence suggests that social mechanisms affect belief formation and the trading dynamics of lottery stocks.

5. Tests Using the Facebook Social Connectedness Index

Using the StockTwits dataset, we have provided tests of our hypothesis that social interactions contribute to both the rise of a lottery stock and investors' attraction to such stocks. This is informative, as StockTwits is the largest finance-specific social media platform, and because the granular, message-level of the data enables us to directly map social interaction activity with market outcomes such as stock returns, aggressive net

buying by retail investors (especially Robinhood investors), disagreement, and aggregate trading volume. Nevertheless, the question arises of how representative the activity and opinions expressed on the platform are of trading decisions in the general retail investor population.

In this section, we complement these tests by using an alternative measure of social interactions that is representative at the population scale. We adopt the Facebook Social Connectedness Index (SCI) as an alternative measure of the extent of investors' social interaction. Introduced by Bailey et al. (2018b), Facebook Social Connectedness Index (SCI) is the friendship probability of Facebook users between two counties.[27] Using this measure, we examine whether the social connectedness of a lottery stock's headquarters location contributes to investor attraction to the stock.

Specifically, we perform tests of the association between social connections and returns to lottery stocks, based on the social connectedness of the stock's headquarters county to investors in the rest of the US. There is extensive evidence that shows that investors are more attentive to nearby firms and are more likely to invest in and trade these firms' stocks.[28] As investors discuss their gains from investing in local stocks, particularly those with lottery-like characteristics, news about these stocks spreads from local investors to those in other counties through word-of-mouth. Consequently, such stocks tend to be extensively discussed among a wider range of investors. As theorized in the aforementioned models (Han, Hirshleifer, and Walden 2022, Pedersen 2022), such a tendency for more intense social interactions regarding the focal stock causes greater investor attraction to stocks that have produced extreme positive returns.

[27] $\text{SCI}_{ij} = \frac{\text{Connections}_{ij}}{\text{Users}_i \cdot \text{Users}_j}$, where Connections_{ij} is the number of Facebook friendship connections between the two counties and Users_i and Users_j represent the number of Facebook users in the corresponding counties. We obtained the 2016 SCI measure from https://data.humdata.org/dataset/social-connectedness-index. This measure has been increasingly used in recent studies as a proxy for real-world social connections owing to its scale and the relative representativeness of its user base. SCI is also relatively stable over time. See, e.g., Bailey et al. (2018b), Bailey et al. (2020), Bailey et al. (2021), and Chetty et al. (2022) on these points.

[28] See, for example, Coval and Moskowitz (1999), Huberman (2001), Ivković and Weisbenner (2005), Ivković and Weisbenner (2007), Hong, Kubik, and Stein (2008), Massa and Simonov (2006), Seasholes and Zhu (2010), and Hong et al. (2014). Additionally, there is a notable local bias in Google searches for firms' stock tickers (Chi and Shanthikumar 2017).

Motivated by these models, we hypothesize that lottery events experienced by stocks in highly connected counties will be associated with higher valuations and lower future returns.

To test this hypothesis, we measure the social connectedness of a stock's headquarters county (SCIH) as the sum of the SCIs between that county and all other US counties (including the headquarters county itself). SCIH therefore serves as a proxy for the strength of social ties between the stock's headquarters county and the rest of the U.S. A higher value of SCIH implies that a lottery stock is more likely to be discussed by a broader set of investors in the social network, thereby triggering a stronger demand for the stock.

We then modify equation (2) by replacing $\mathbb{1}_{mssg}$ with SCIH and estimate the model for the same sample period. Table 8 presents the results and shows that the coefficients of MAXRET·SCHI are consistently negative and significant in predicting one-, two-, and three-month-ahead stock returns. The coefficient in column 1 suggests that a one-standard-deviation increase in MAXRET is associated with more negative one-month-ahead returns, by 85 basis points, for stocks in the highest SCIH decile compared to those in the lowest decile.[29] The corresponding values are 114 and 121 basis points for two- and three-month-ahead returns, respectively.

The result is consistent with our hypothesis that lottery stocks headquartered in socially connected areas are more likely to be overvalued and, therefore, more likely to exhibit lower subsequent returns.

6. Further Analyses

In this section, we complement our main results with additional analyses that strengthen identification, rule out alternative explanations, and confirm robustness. Specifically, we first exploit exogenous shocks to social interactions to support causality. We then com-

[29]This is calculated as the product of the standard deviation of MAXRET (9.32%), the spread in SCIH between the top and bottom SCIH deciles (3.95), and the slope coefficient of MAXRET·SCIH (−0.023) from column 1.

pare social interactions with traditional retail and institutional attention measures to distinguish their roles. Finally, we demonstrate robustness to alternative definitions of lottery-like stocks and events, as well as to specifications that include additional controls for the information environment and arbitrage costs.

6.1. Exogenous Shocks to Social Interactions

We first address identification by using exogenous regulatory shocks to isolate the causal effect of social interactions on lottery demand. Specifically, we exploit the staggered implementation of state-level social media privacy laws as an exogenous shock to social media activity.

Beginning in 2012, U.S. states enacted legislation to protect employees' and job applicants' personal social media accounts from employer intrusion. These laws generally prohibit employers from requesting or requiring disclosure of passwords to personal accounts, demanding access to such accounts, or taking adverse employment actions based on an individual's refusal to provide such access. The laws were adopted in response to growing concerns about employer overreach into employees' private online activities.

These privacy protections mitigate employees' concerns about employer surveillance of their social media activities. Prior to these laws, employees might have self-censored their online content, worrying about potential monitoring by employers. By reducing these concerns, the laws create an environment in which individuals feel more secure engaging in open online discussions. This increased freedom of expression plausibly stimulates more vibrant social interactions, making the laws a relevant instrument for social media activity.

Online Appendix Table S.III presents the implementation details of these laws across the U.S. states.[30] In our sample period, 26 states have implemented these laws, affecting 44% of the firms in our sample at some point. The adoption of these laws is well-staggered over time, providing substantial cross-sectional and time-series variations in users' free-

[30]We use data from the Seyfarth Shaw's Social Media Privacy Legislation Survey (https://www.seyfarth.com/a/web/7266/131317SocialMediaSurveyM13.pdf) supplemented with additional research to identify implementation dates and statutory citations.

dom of online expressions and therefore variations in the intensity of their social media activities.

We employ a two-stage least squares (2SLS) approach, using these privacy law implementations as an instrument for StockTwits messaging activity. Our instrumental variable, SMP, is an indicator that equals one if a firm is headquartered in a state with an active social media privacy law in a given year, and zero otherwise. This instrument captures a widespread regulatory shift rather than a localized event.

Using this instrument, we examine whether intense StockTwits discussions lead to lottery events, complementing the OLS results reported in Table 3. Table 9 presents the 2SLS results. In the full sample of stock-day obserations (columns 1−2), the first-stage shows that SMP significantly increases StockTwits message activity, with a coefficient of 0.014 (t-statistic = 2.87). The first-stage F-statistic of 1,839.7 exceeds conventional thresholds for weak instruments, confirming the relevance of our instrument. The second-stage result (column 2) provides causal evidence for our main hypothesis: instrumented social media activity ($\mathbb{1}_{mssg}$) significantly predicts lottery events, with a coefficient of 2.790 (t-statistic = 2.54).[31]

Columns 3 and 4 report results for the lottery-event day sample. The first stage (column 3) again shows that SMP increases StockTwits activity, although the instrument is weaker, with an F-statistic of 12.7. The second stage (column 4) yields an insignificant coefficient on instrumented social media activity ($\mathbb{1}_{mssg}$), likely reflecting the much smaller sample of event days. Nevertheless, the positive sign is consistent with our hypothesis that social media activity amplifies lottery-like characteristics.

Overall, these results suggest that exogenous increases in social media activity causally contribute to both the emergence of lottery stocks and their return dynamics. This causal evidence mitigates concerns that the observed link between social media activity and lottery demand effect is driven by reverse causality or omitted variables.

[31]Notably, the IV coefficient (2.790) is substantially larger than the OLS coefficient (0.272) reported in Table 3. This pattern is well-documented in the literature (e.g., Card 2001; Jiang 2017) and can arise due to the presence of omitted variable bias, endogeneity, or weak instruments. In our setting, the high F-statistic suggests weak instruments are not the concern).

6.2. Social Interactions versus Traditional Attention Measures

We next disentangle the role of social interactions from general investor attention by introducing established attention proxies. A potential concern is that our StockTwits message volume may primarily reflect general investor attention rather than genuine social interactions. To address this issue, we control for two established attention measures: abnormal Google search volume (ASV), which primarily captures retail investor attention, and Bloomberg daily maximum readership (DMR), which represents institutional investor attention.[32]

We incorporate ASV and DMR into the regression specification given in equation (2), where ASV and DMR are measured as the average daily over the window $[-11, 0]$ relative to a MAX event. The corresponding results are reported in Table 10. The interaction between MAX returns and StockTwits activity (MAX $\cdot \mathbb{1}_{mssg}$) remains highly significant across all horizons, with coefficients ranging from -0.162 to -0.263, which are comparable to our baseline results presented in Table 8.

Interestingly, we find that the two attention measures exhibit distinct effects in modulating the MAX-return relationship. The MAX·ASV interaction is insignificant, suggesting that retail attention captured through Google searches does not impact the return reversals of lottery stocks. In contrast, the MAX·DMR interaction is positive and statistically significant, with coefficients ranging from 0.080 to 0.141, indicating that institutional attention attenuates the lottery demand effect. This finding is consistent with institutional investors exerting a stabilizing influence on lottery stock prices.

[32]We construct the two attention measures following Da, Engelberg, and Gao (2011) and Ben-Rephael, Da, and Israelsen (2017). The first measure, abnormal Google search volume (ASV), is defined as the difference between Google's search popularity score for a given ticker symbol and its one-year lagged mean, scaled by the mean. To avoid spillover from recent events, we exclude the most recent month when computing the mean. The ASV sample covers the period from July 2004 to December 2022. The second measure, Bloomberg daily maximum readership (DMR), is available from February 2010 to December 2022. The Bloomberg data track hourly user activity (including search and readership) for each stock relative to the same stock's activity over the prior 30 days. DMR is an ordinal measure, taking the value of zero, one, two, three, or four depending on whether the day's maximum hourly activity falls below the 80th percentile, between the 80th and 90th, between the 90th and 94th, between the 94th and 96th, or above the 96th percentile of the stock's historical distribution.

We extend this analysis to the prediction of lottery events in Panel B. Even after controlling for the averages of ASV and DMR in the pre-event window $[-11, -1]$, the coefficient on $\mathbb{1}_{mssg}^{preEvent}$ remains highly significant at 0.247 for predicting lottery event probability and 3.598 for predicting lottery event returns. The magnitudes are similar to those reported in Table 3. Both of the conventional attention measures positively predict lottery events, but their inclusion does not diminish the role of StockTwits messaging.

Together, these results reinforce our interpretation that social media activity captures a distinct dimension of investor behavior that differs from individualized information-seeking as measured by retail search volume or institutional readership.

6.3. Alternative Lottery Measure

To ensure that our findings are not specific to the MAX measure, we construct a lottery index, LTRY, adapted from Kumar (2009). LTRY captures three key features that make stocks lottery-like from the perspective of retail investors: high idiosyncratic volatility, high idiosyncratic skewness, and low nominal price.[33]

Appendix Table A5 examines the degree of overpricing of lottery stocks. Similar to Table 2, the interaction between LTRY and StockTwits activity (LTRY$\cdot\mathbb{1}_{mssg}$) significantly predicts negative future returns across all horizons, with coefficients ranging from -0.127 to -0.365. These magnitudes are economically meaningful: a one-decile-rank increase in LTRY is associated with an additional 12.7 to 36.5 basis-point decrease in future returns over one to three months for stocks with high StockTwits activities. Also consistent with Table 2, the standalone lottery indicator is insignificant, reinforcing the idea that social interaction are a key factor behind the underperformance of lottery stocks in our sample.

[33]For each month, we sort stocks into 50 bins by price per share (PRC) in descending order, such that stocks in the lowest bin (i.e., a PRC portfolio rank of 1) have the highest price per share and those in the highest bin (i.e., a PRC portfolio rank of 50) have the lowest price per share. We also independently sort stocks into 50 bins by idiosyncratic volatility (IVOL) and idiosyncratic skewness (ISKEW) in ascending order. IVOL and ISKEW are, respectively, the standard deviation and the skewness of residuals from the time series regression of daily stock returns against the daily market, size, and book-to-market factors in a month. The three ranks are summed to create a composite lottery index (SumRank). By construction, the lottery index, SumRank, has an integer value in the range from 3 to 150, and higher values indicate stronger lottery-like characteristics. Finally, we define LTRY as decile rank of SumRank, with the highest decile (LTRY = 10) representing the most lottery-like stocks.

Table A6 in the appendix further examines whether social media activity predicts lottery events defined based on LTRY. We classify a lottery event as a stock falling into the top LTRY decile in a given month and define $\mathbb{1}_{mssg}^{preMonth}$ as equal to one if lagged monthly message volume ranks in the top 10% cross-sectionally. The results show that $\mathbb{1}_{mssg}^{preMonth}$ significantly predicts the probability of achieving lottery status (coefficients of 0.022 to 0.029). The economic magnitude is sizable: high social media activity increases the probability of lottery classification by 2.9 percentage points, a 29% increase relative to the baseline monthly probability of 10%.

Taken together, these results demonstrate that our main finding, that social interactions contribute to the emergence of lottery stocks, is robust to the alternative lottery measure of Kumar (2009). This mitigates the concern that our results are driven by a particular operational definition of lottery characteristics.

6.4. Additional Robustness Checks

This subsection provides additional robustness checks. We show below that our main findings remain similar with alternative windows over which we measure message activity, after controlling for information environment, information supply, arbitrage costs, and for alternative definitions of lottery events.

Alternative message windows We construct versions of Table 2 using alternative definitions of StockTwits message indicator, $\mathbb{1}_{mssg}$, that are measured with different event windows: $[-5, 0]$, $[-3, 0]$, and $[-1, 0]$. The results are presented in Table A7 of the appendix. The slope coefficients of MAXRET$\cdot\mathbb{1}_{mssg}$, the primary variable of interest, are negative and statistically significant, ranging from -0.103 to -0.192 for the $[-5, 0]$ window, from -0.101 to -0.155 for the $[-3, 0]$ window, and from -0.083 to -0.136 for the $[-1, 0]$ window. These results are similar to those reported in Table 2.

Information environment and information supply Our findings so far suggest that social interactions help direct investor attention to lottery stocks and therefore contribute

to lottery demand. We next show that this effect is not driven by a stock's information environment as measured by the number of analysts covering a stock (e.g., Hou and Moskowitz 2005; Hong, Torous, and Valkanov 2007) nor by information supply, as captured by news coverage.

We obtain analyst coverage data from I/B/E/S for 1976–2022 and define $CVRG$ as the natural logarithm of one plus the number of analysts covering a firm. For information supply, we use the number of relevant news reports ($NEWS$) from Ravenpack. We estimate equation (2) with $CVRG$ or $NEWS$ as additional controls and present the results in Appendix Table A8, Panels A and B. The average slope coefficients on the MAXRET and $\mathbb{1}_{mssg}$ interaction remain significantly negative, with magnitudes similar to those in Table 2. This evidence suggests that our main finding on the relation between social interactions and lottery stock returns is not driven by information environment or information supply.

Arbitrage costs High arbitrage barriers discourage investors from trading, thereby allowing mispricing to persist (see, for example, Shleifer and Vishny 1997). We therefore consider the extent to which our result in Table 2 is driven by arbitrage costs.

We measure arbitrage costs in two ways: idiosyncratic volatility (IVOL) and an arbitrage cost index (COST).[34] We then estimate equation (2) with the inclusion of IVOL or COST as an additional control variable and present the results in Panels C and D of Table A8 in the appendix, respectively. We find that the average slope coefficients on MAXRET$\cdot\mathbb{1}_{mssg}$ are significantly negative and remain similar to the corresponding results reported in Table 2. Hence, the results suggest that arbitrage costs do not explain our main finding.

[34]Idiosyncratic volatility (IVOL) has been utilized as a proxy for risk by arbitrageurs seeking to exploit mispricing (e.g., Pontiff 2006; Stambaugh, Yu, and Yuan 2015). Following Ang et al. (2006), IVOL is estimated from stock-level time-series regressions of daily excess stock returns (with a minimum of 15 daily observations) against the daily Fama-French market, size, and book-to-market factors during the portfolio formation month. We measure COST following Stambaugh, Yu, and Yuan (2012, 2015): for each month, stocks are independently sorted into decile portfolios based on institutional ownership, market capitalization, analyst coverage, illiquidity, idiosyncratic volatility, and the number of months a stock is present on the CRSP database. COST is the arithmetic average of the decile ranks of the six stock characteristics, with the requirement that a minimum of three characteristics available.

Alternative lottery event definition So far, our definition of a lottery event on a given day depends on a stock experiencing a MAX event on that day, where MAX event occurs when a stock's return on that day is the highest over a trailing 21-trading-day window. As explained in Subsection 2.1, we use this definition of MAX events because it is the natural generalization of the specification used in previous papers that examine lottery effects (i.e., BCW) while only using backward-looking data to construct the predictor. This allows us to compare our results with the previous literature and evaluate the role of social interaction variables in modulating the relation between MAXRET and future returns (Table 2).

A potential concern with this measure is that it is possible that omitted shocks to the stock over the pre-event period may influence both MAXRET and message volume and therefore drive some of our findings. To address this possible source of endogeneity, we perform robustness checks using two alternative definitions of lottery event.

The first is a path-independent lottery event definition, with a lottery event occurring if a stock's return on a day ranks in the top 10% of the cross-sectional distribution of the highest daily returns in the trailing 21 trading days. This definition does not depend on the stock's own past returns, thus alleviating the endogeneity concern mentioned in the paragraph above.

The second alternative definition introduces a 10-day gap between a stock's daily return and the window of past returns. Specifically, a MAX event is defined as occurring when a stock's return on day t is equal to or greater than its returns over the $[t-32, t-12]$ window. A stock experiences a lottery event if it has a MAX event and the MAX-day return (MAXRET) falls within the top 10% of the cross-sectional distribution. This definition ensures the past return window does not overlap with the window over which we measure pre-lottery event message activity.

We reproduce our key tests, Tables 2—5, using the alternative definitions and present the results in Online Appendix Tables S.IV—S.VII. These tables show patterns that are

consistent with the corresponding tables presented in the main text. This indicates that our main findings are robust to the alternative lottery definitions.

7. Conclusion

We explore here the hypothesis that social interactions promote investor attention to and demand for lottery stocks, and thereby the formation and amplification of lottery stock bubbles. Based on a variety of tests using data from two types of social media and multiple outcome variables (extreme return realizations, subsequent stock returns and earning announcement returns, retail buying, particularly by Robinhood users, investor disagreement proxies, and trading volume), our evidence lends support to recent social finance models of bubbles.

While most previous studies take a stock's lotteriness as given, we show that social interactions, measured by StockTwits message volume, predict a higher likelihood of a stock becoming a lottery stock. Exploiting the staggered adoption of state-level social media privacy laws as an exogenous shock, we establish that the effect of StockTwits activity on both the likelihood of a lottery event and, conditional on the event, the magnitude of lottery returns is causal. Importantly, these effects are distinct from those of traditional attention proxies such as Google search volume and Bloomberg readership by institutional investors.

We also provide new insights into the source of return predictive power of lottery characteristics documented in prior research. In our sample, the predictive power of both the maximum recent daily return and the composite lottery index is concentrated in stocks with high StockTwits message volume. This indicates that social interactions modulate the relationship between a stock's lottery characteristic and its subsequent returns and are an important source of the cross-sectional predictive power of the lottery characteristic.

We complement the StockTwits-based tests using Facebook's social connectivity data between users' county locations. We find that lottery stocks of firms headquartered in more socially connected counties experience more negative subsequent-month returns than

for firms in less-connected counties. This suggests that the social connectivity of investors in a stock's headquarter county contributes to the overvaluation of lotteriness.

Overall, these findings lend support for models in which social interactions promote investor attraction to lottery stocks. This raises the interesting possibility that social mechanisms underlie other trading patterns and pricing anomalies. Such mechanisms are likely to be particularly important given increased online social interactions and the proliferation of zero-commission trading platforms.

References

Agarwal, V., L. Jiang, and Q. Wen. 2021. Why do mutual funds hold lottery stocks? *Journal of Financial and Quantitative Analysis* 57:825–56.

Ahern, K. R. 2017. Information networks: Evidence from illegal insider trading tips. *Journal of Financial Economics* 125:26–47.

Amihud, Y. 2002. Illiquidity and stock returns: Cross-section and time series effects. *Journal of Financial Markets* 5:31–56.

An, L., H. Wang, J. Wang, and J. Yu. 2020. Lottery-related anomalies: The role of reference-dependent preferences. *Management Science* 66:473–501.

Andrei, D., and M. Hasler. 2015. Investor attention and stock market volatility. *Review of Financial Studies* 28:33–72.

Ang, A., R. Hodrick, Y. Xing, and X. Zhang. 2006. The cross-section of volatility and expected returns. *Journal of Finance* 61:259–99.

Annaerta, J., M. DeCeustera, and K. Verstegena. 2013. Are extreme returns priced in the stock market? European evidence. *Journal of Banking and Finance* 37:3401–11.

Antweiler, W., and M. Z. Frank. 2004. Is all that talk just noise? The information content of internet stock message boards. *Journal of Finance* 59:1259–94.

Atilgan, Y., T. G. Bali, K. O. Demirtas, , and A. D. Gunaydin. 2020. Left-tail momentum: Underreaction to bad news, costly arbitrage and equity returns. *Journal of Financial Economics* 135:725–75.

Bailey, M., R. Cao, T. Kuchler, and J. Stroebel. 2018a. The economic effects of social networks: Evidence from the housing market. *Journal of Political Economy* 126:2224–76.

Bailey, M., R. Cao, T. Kuchler, J. Stroebel, and A. Wong. 2018b. Social connectedness: Measurement, determinants, and effects. *Journal of Economic Perspective* 32:259–80.

Bailey, M., A. Gupta, S. Hillenbrand, T. Kuchler, R. Richmond, and J. Stroebel. 2021. International trade and social connectedness. *Journal of International Economics* 129:103418.

Bailey, M., D. Johnston, T. Kuchler, D. Russel, B. State, and J. Stroebel. 2020. The determinants of social connectedness in Europe. In S. Aref, K. Bontcheva, M. Braghieri, F. Dignum, F. Giannotti, F. Grisolia, and D. Pedreschi, eds., *Social Informatics*, 1–14. Cham: Springer International Publishing.

Bali, T. G., S. Brown, S. Murray, and Y. Tang. 2017. A lottery demand-based explanation of the beta anomaly. *Journal of Financial and Quantitative Analysis* 52:2369–97.

Bali, T. G., S. Brown, and Y. Tang. 2017. Is economic uncertainty priced in the cross-section of stock returns? *Journal of Financial Economics* 126:471–89.

Bali, T. G., N. Cakici, and R. Whitelaw. 2011. Maxing out: Stocks as lotteries and the cross-section of expected returns. *Journal of Financial Economics* 99:427–46.

Bali, T. G., A. D. Gunaydin, T. Jansson, and Y. Karabulut. 2023. Do the rich gamble in the stock market? low risk anomalies and wealthy households. *Journal of Financial Economics* 150:103715.

Banerjee, A. V. 1992. A simple model of herd behavior. *Quarterly Journal of Economics* 107:797–817.

Barber, B., and T. Odean. 2008. All that glitters: The effect of attention on the buying behavior of individual and institutional investors. *Review of Financial Studies* 21:785–818.

Barber, B. M., X. Huang, T. Odean, and C. Schwarz. 2022. Attention-induced trading and returns: Evidence from robinhood users. *Journal of Finance* 77:3141–90.

———. 2024. A (sub)penny for your thoughts: Tracking retail investor activity in taq. *Journal of Finance* 79:2403–27.

Barberis, N., and M. Huang. 2008. Stocks as lotteries: The implications of probability weighting for security prices. *American Economic Review* 98:2066–100.

Barberis, N., A. Mukherjee, and B. Wang. 2016. Prospect theory and stock returns: An empirical test. *Review of Financial Studies* 29:3068–107.

Ben-Rephael, A., Z. Da, and R. D. Israelsen. 2017. It depends on where you search: Institutional investor attention and underreaction to news. *Review of Financial Studies* 30:3009–47. doi: 10.1093/rfs/hhx036.

Bikhchandani, S., D. Hirshleifer, and I. Welch. 1992. A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of Political Economy* 100:992–1026.

Boehmer, E., C. M. Jones, X. Zhang, and X. Zhang. 2021. Tracking retail investor activity. *Journal of Finance* 76:2249–305.

Bordalo, P., N. Gennaioli, and A. Shleifer. 2012. Salience theory of choice under risk. *Quarterly Journal of Economics* 127:1243–85.

———. 2013. Salience and asset prices. *American Economic Review* 103:623–8.

Boyer, B., T. Mitton, and K. Vorkink. 2010. Expected idiosyncratic skewness. *Review of Financial Studies* 23:169–202.

Boyer, B., and K. Vorkink. 2014. Stock options as lotteries. *Journal of Finance* 69:1485–528.

Brown, J. R., Z. Ivković, P. A. Smith, and S. Weisbenner. 2008. Neighbors matter: Causal community effects and stock market participation. *Journal of Finance* 63:1509–31.

Brunnermeier, M. K., and J. A. Parker. 2005. Optimal expectations. *American Economic Review* 95:1092–118.

Buffett, W. 2024. 2024 letter to Berkshire Hathaway shareholders.

Burnside, C., M. Eichenbaum, and S. Rebelo. 2016. Understanding booms and busts in housing markets. *Journal of Political Economy* 124:1088–147.

Card, D. 2001. Estimating the return to schooling: Progress on some persistent econometric problems. *Econometrica* 69:1127–60.

Carpenter, J., F. Lu, and R. Whitelaw. 2021. The real value of china's stock market. *Journal of Financial Economics* 139:679–96.

Chen, H., P. De, Y. J. Hu, and B.-H. Hwang. 2014. Wisdom of crowds: The value of stock opinions transmitted through social media. *Review of Financial Studies* 27:1367–403.

Chen, H. C., and B.-H. Hwang. 2022. Listening in on investors' thoughts and conversations. *Journal of Financial Economics* 145:426–44.

Chetty, R., M. O. Jackson, T. Kuchler, J. Stroebel, N. Hendren, R. B. Fluegge, S. Gong, F. Gonzalez, A. Grondin, M. Jacob, D. Johnston, M. Koenen, E. Laguna-Muggenburg, F. Mudekereza, T. Rutter, N. Thor, W. Townsend, R. Zhang, M. Bailey, P. Barberá, M. Bhole, and N. Wernerfelt. 2022. Social capital I: Measurement and associations with economic mobility. *Nature* 608:108–21.

Chi, S. S., and D. M. Shanthikumar. 2017. Local bias in Google search and the market response around earnings announcements. *Accounting Review* 92:115–43.

Cohen, L., A. Frazzini, and C. J. Malloy. 2008. The small world of investing: Board connections and mutual fund returns. *Journal of Political Economy* 116:951–79.

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

Cookson, J. A., R. Lu, W. Mullins, and M. Niessner. 2024. The social signal. *Journal of Financial Economics* 158:103870.

Cookson, J. A., W. Mullins, and M. Niessner. 2024. Social media and finance. *Oxford Research Encyclopedia of Economics and Finance* Forthcoming.

Cookson, J. A., and M. Niessner. 2020. Why don't we agree? Evidence from a social network of investors. *Journal of Finance* 75:173–228.

———. 2023. Investor disagreement: Daily measures from social media. Working Paper.

Coval, J. D., and T. J. Moskowitz. 1999. Home bias at home: Local equity preference in domestic portfolios. *Journal of Finance* 54:2045–73.

Crawford, S. S., W. R. Gray, and A. E. Kern. 2017. Why do fund managers identify and share profitable ideas? *Journal of Financial and Quantitative Analysis* 52:1903–26.

Da, Z., J. Engelberg, and P. Gao. 2011. In search of attention. *Journal of Finance* 66:1461–99.

Daniel, K. D., M. Grinblatt, S. Titman, and R. Wermers. 1997. Measuring mutual fund performance with characteristic-based benchmarks. *Journal of Finance* 52:1035–58.

DeMarzo, P. M., R. Kaniel, and I. Kremer. 2008. Relative wealth concerns and financial bubbles. *Review of Financial Studies* 21:19–50.

DeMarzo, P. M., D. Vayanos, and J. Zwiebel. 2003. Persuasion bias, social influence, and unidimensional opinions. *Quarterly Journal of Economics* 118:909–68.

Devlin, J., M.-W. Chang, K. Lee, and K. Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In J. Burstein, C. Doran, and T. Solorio, eds., *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, 4171–86. Minneapolis, Minnesota: Association for Computational Linguistics.

Duflo, E., and E. Saez. 2002. Participation and investment decisions in a retirement plan: The influence of colleagues' choices. *Journal of Public Economics* 85:121–48.

———. 2003. The role of information and social interactions in retirement plan decisions: Evidence from a randomized experiment. *Quarterly Journal of Economics* 118:815–42.

Eaton, G. W., T. C. Green, B. S. Roseman, and Y. Wu. 2022. Retail trader sophistication and stock market quality: Evidence from brokerage outages. *Journal of Financial Economics* 146:502–28.

Engelberg, J., D. McLean, and J. Pontiff. 2018. Anomalies and news. *Journal of Finance* 73:1971–2001.

Fama, E. F., and K. R. French. 1992. The cross-section of expected stock returns. *Journal of Finance* 46:427–66.

———. 2015. A five-factor asset pricing model. *Journal of Financial Economics* 116:1–22.

Farrell, M., T. C. Green, R. Jame, and S. Markov. 2022. The democratization of investment research and the informativeness of retail investor trading. *Journal of Financial Economics* 145:616–41.

Frazzini, A. 2006. The disposition effect and under-reaction to news. *Journal of Finance* 61:2017–46.

Gao, X., and J. R. Ritter. 2010. The marketing of seasoned equity offerings. *Journal of Financial Economics* 97:33–52.

Giannini, R., P. Irvine, and T. Shu. 2019. The convergence and divergence of investors' opinions around earnings news: Evidence from a social network. *Journal of Financial Markets* 42:94–120.

Gomez, J.-P., R. Priestley, and F. Zapatero. 2009. Implications of keeping-up-with-the-joneses behavior for the equilibrium cross section of stock returns: International evidence. *Journal of Finance* 64:2703–37.

Green, T. C., and B.-H. Hwang. 2012. IPOs as lotteries: Skewness preference and first-day returns. *Management Science* 58:432–44.

Green, T. C., and R. Jame. 2024. Retail trading frenzies and real investment. Working Paper.

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

Han, B., and A. Kumar. 2013. Speculative retail trading and asset prices. *Journal of Financial and Quantitative Analysis* 377–404.

Han, B., and L. Yang. 2013. Social networks, information acquisition, and asset prices. *Management Science* 59:1444–57.

Harris, M., and A. Raviv. 1993. Differences of opinion make a horse race. *Review of Financial Studies* 6:475–506.

Harvey, C. R., and A. Siddique. 2000. Conditional skewness in asset pricing tests. *Journal of Finance* 55:1263–95.

Heimer, R. 2016. Peer pressure: Social interaction and the disposition effect. *Review of Financial Studies* 29:3177–209.

Hirshleifer, D. 2020. Presidential address: Social transmission bias in economics and finance. *Journal of Finance* 75:1779–831.

Hirshleifer, D., L. Peng, and Q. Wang. 2024. News diffusion in social networks and stock market reactions. *Review of Financial Studies* 883–937.

Hong, H., W. Jiang, N. Wang, and B. Zhao. 2014. Trading for status. *Review of Financial Studies* 27:3171–212.

Hong, H., J. Kubik, and J. Stein. 2004. Social interaction and stock market participation. *Journal of Finance* 59:137–63.

———. 2005. Thy neighbor's portfolio: Word-of-mouth effects in the holdings and trades of money managers. *Journal of Finance* 60:2801–24.

———. 2008. The only game in town: Stock-price consequences of local bias. *Journal of Financial Economics* 90:20–37.

Hong, H., W. Torous, and R. Valkanov. 2007. Do industries lead the stock market? *Journal of Financial Economics* 83:367–96.

Hong, H., and J. Xu. 2019. Inferring latent social networks from stock holdings. *Journal of Financial Economics* 131:323–44.

Hou, K., and T. Moskowitz. 2005. Market frictions, price delay, and the cross-section of expected returns. *Review of Financial Studies* 18:981–1020.

Hou, K., C. Xue, and L. Zhang. 2015. Digesting anomalies: An investment approach. *Review of Financial Studies* 28:650–705.

Hu, D., C. M. Jones, S. Li, V. Zhang, and X. Zhang. 2021. The rise of reddit: How social media affects belief formation and price discovery. Working paper.

Huang, S., B.-H. Hwang, and D. Lou. 2021. The rate of communication. *Journal of Financial Economics* 141:533–50.

Huberman, G. 2001. Familiarity breeds investment. *Review of Financial Studies* 14:659–80.

Ivković, Z., and S. Weisbenner. 2005. Local does as local is: Information content of the geography of individual investors' common stock investments. *Journal of Finance* 60:267–306.

———. 2007. Information diffusion effects in individual investors' common stock purchases: Covet thy neighbors' investment choices. *Review of Financial Studies* 20:1327–57.

Jackson, M. O., S. Malladi, and D. McAdams. 2024. The impossibility of social learning with noisy communication: An identification failure. Working Paper.

Jegadeesh, N. 1990. Evidence of predictable behavior of security returns. *Journal of Finance* 45:881–98.

Jegadeesh, N., and S. Titman. 1993. Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance* 48:65–91.

Jiang, W. 2017. Have instrumental variables brought us closer to the truth. *Review of Corporate Finance Studies* 6:127–40.

Kandel, E., and N. D. Pearson. 1995. Differential interpretation of public signals and trade in speculative markets. *Journal of Political Economy* 103:831–72.

Kaniel, R., S. Liu, G. Saar, and S. Titman. 2012. Individual investor trading and return patterns around earnings announcements. *Journal of Finance* 67:639–80.

Kaustia, M., and S. Knüpfer. 2012. Peer performance and stock market entry. *Journal of Financial Economics* 104:321–38.

Kelley, E. K., and P. C. Tetlock. 2013. How wise are crowds? Insights from retail orders and stock returns. *Journal of Finance* 68:1229–65.

———. 2017. Retail short selling and stock prices. *Review of Financial Studies* 30:801–34.

Kelly, M., and C. O. O'Grada. 2000. Market contagion: Evidence from the panics of 1854 and 1857. *American Economic Review* 90:1110–24.

Kim, O., and R. E. Verrecchia. 1991. Trading volume and price reactions to public announcements. *Journal of Accounting Research* 29:302–21.

Kuchler, T., and J. Stroebel. 2021. Social finance. *Annual Review of Financial Economics* 13:37–55.

Kumar, A. 2009. Who gambles in the stock market? *Journal of Finance* 64:1889–933.

Kumar, A., J. K. Page, and O. G. Spalt. 2011. Religious beliefs, gambling attitudes, and financial market outcomes. *Journal of Financial Economics* 102:671–708.

———. 2016. Gambling and comovement. *Journal of Financial and Quantitative Analysis* 51:85–111.

Lin, T.-C., and X. Liu. 2018. Skewness, individual investor preference, and the cross-section of stock returns. *Review of Finance* 22:1841–76.

Liu, B., H. Wang, J. Yu, and S. Zhao. 2020. Time-varying demand for lottery: Speculation ahead of earnings announcements. *Journal of Financial Economics* 138:789–817.

Lu, J., N.-T. Yang, K.-Y. Ho, and K.-C. Ko. 2022. Lottery demand and the asset growth anomaly. *Finance Research Letters* 48:102988–.

Martin, K., and R. Wigglesworth. 2021. Rise of the retail army: The amateur traders transforming markets. *Financial Times* .

Massa, M., and A. Simonov. 2006. Hedging, familiarity and portfolio choice. *Review of Financial Studies* 19:633–85.

Maturana, G., and J. Nickerson. 2018. Teachers teaching teachers: The role of workplace peer effects in financial decisions. *Review of Financial Studies* 32:3920–57.

Mitton, T., K. Vorkink, and I. Wright. 2018. Neighborhood effects on speculative behavior. *Journal of Economic Behavior & Organization* 151:42–61.

Ouimet, P. P., and G. A. Tate. 2020. Learning from coworkers: Peer effects on individual investment decisions. *Journal of Finance* 75:133–72.

Ozik, G., R. Sadka, and S. Shen. 2021. Flattening the illiquidity curve: Retail trading during the COVID-19 lockdown. *Journal of Financial and Quantitative Analysis* 56:2356–88.

Pedersen, L. H. 2022. Game on: Social networks and markets. *Journal of Financial Economics* 146:1097–119.

Petersen, M. A. 2009. Estimating standard errors in finance panel data sets: Comparing approaches. *Review of Financial Studies* 22:435–80.

Pontiff, J. 2006. Costly arbitrage and the myth of idiosyncratic risk. *Journal of Accounting and Economics* 42:35–52.

Pool, V. K., N. Stoffman, and S. E. Yonker. 2015. The people in your neighborhood: Social interactions and mutual fund portfolios. *Journal of Finance* 70:2679–732.

Rao, P. 2023. Charted: U.s. retail investor inflows (2014–2023). *Visual Capitalist*.

Roussanov, N. 2010. Diversification and its discontents: Idiosyncratic and entrepreneurial risk in the quest for social status. *Journal of Finance* 65:1755–88.

Scheinkman, J. A., and W. Xiong. 2003. Overconfidence, short-sale constraints, and bubbles. *Journal of Political Economy* 111:1183–219.

Seasholes, M. S., and N. Zhu. 2010. Individual investors and local bias. *Journal of Finance* 65:1987–2010.

Shiller, R. J. 2000. Measuring bubble expectations and investor confidence. *Journal of Psychology and Financial Markets* 1:49–60.

Shive, S. 2010. An epidemic model of investor behavior. *Journal of Financial and Quantitative Analysis* 45:169–98.

Shleifer, A., and R. Vishny. 1997. The limits of arbitrage. *Journal of Finance* 52:35–55.

Shumway, T. 1997. The delisting bias in crsp data. *Journal of Finance* 52:327–40.

Stambaugh, R. F., J. Yu, and Y. Yuan. 2012. The short of it: Investor sentiment and anomalies. *Journal of Financial Economics* 104:288–302.

———. 2015. Arbitrage asymmetry and the idiosyncratic volatility puzzle. *Journal of Finance* 70:1903–48.

Walkshausl, C. 2014. The max effect: European evidence. *Journal of Banking and Finance* 42:1–10.

Wang, H., J. Yan, and J. Yu. 2017. Reference-dependent preferences and the risk-return tradeoff. *Journal of Financial Economics* 123:395–414.

Welch, I. 2022. The wisdom of the robinhood crowd. *Journal of Finance* 77:1489–527.

Yuan, Y. 2015. Market-wide attention, trading, and stock returns. *Market-Wide Attention, Trading, and Stock Returns* 116:548–64.

Zhong, A., and P. Gray. 2016. The max effect: An exploration of risk and mispricing explanations. *Journal of Banking and Finance* 65:76–90.

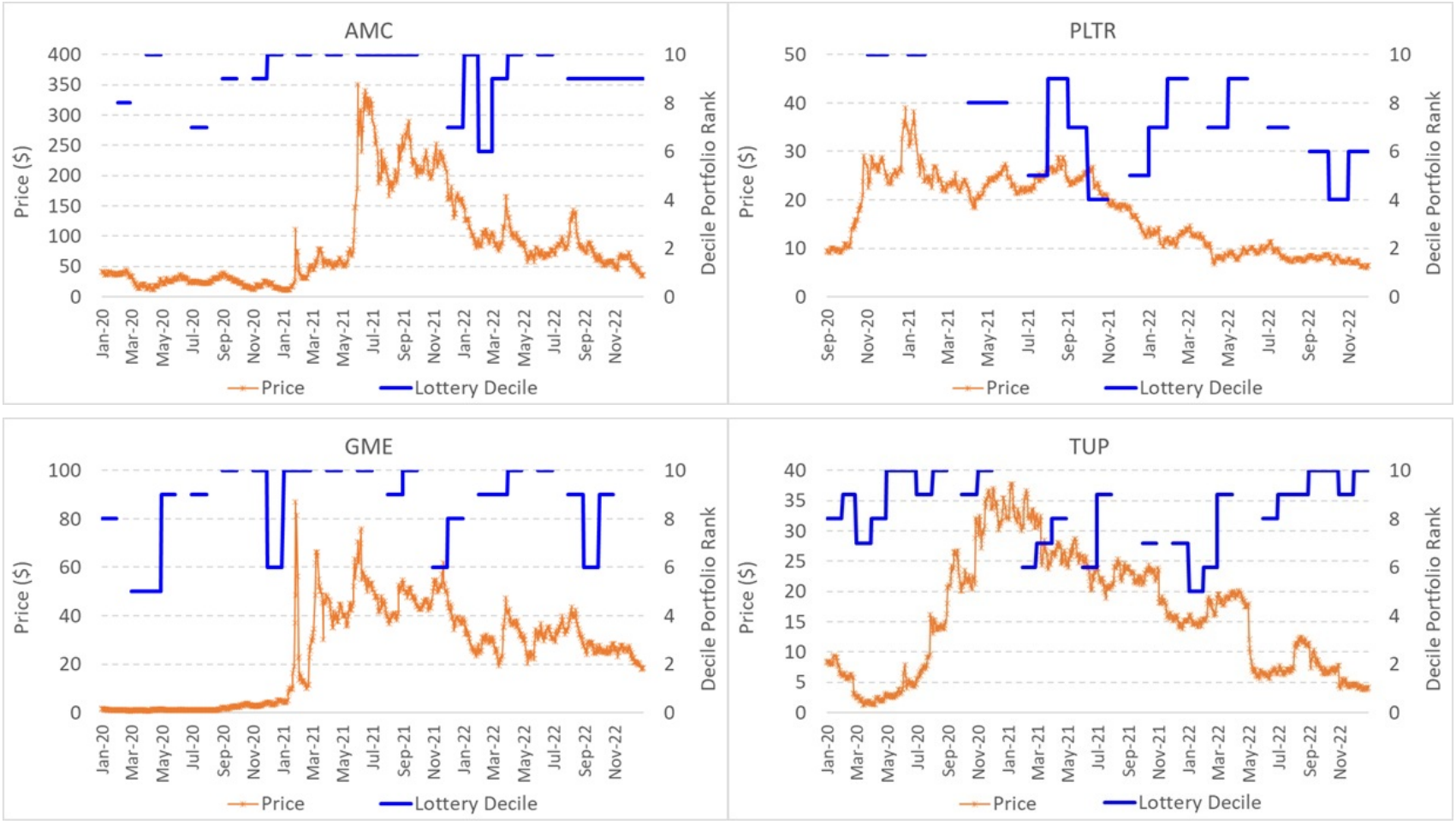


Figure 1. Meme Stocks and Their Lottery-Characteristics This figure shows the prices and lottery decile rankings for the following meme stocks: AMC, GameStop (GME), Palantir (PLTR), and Tupperware (TUP), covering the period from 2020 to 2022. A stock experiences a MAX event on a given day if its return on that day is the highest over a trailing 21-trading-day window. MAXRET refers to the return on the MAX event day and captures a stock's lotteriness. Each month, stocks are sorted into decile portfolios based on MAXRET, with Deciles 10 representing the highest returns and Decile 1 the lowest. If a month has multiple events, the month's highest daily return is assigned as MAXRET.

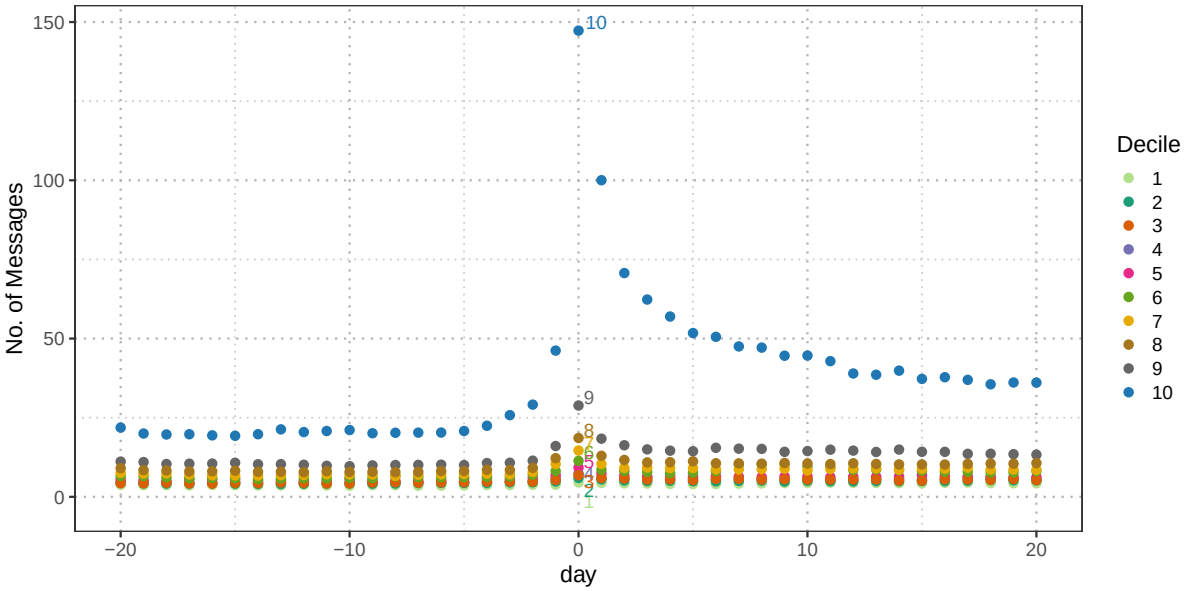


Figure 2. StockTwits Activity around MAX Events This figure shows the average StockTwits message volume around MAX event days (day 0) for MAX decile portfolios. A stock experiences a MAX event on a given day if its return on that day is the highest over a trailing 21-trading day window. Decile 10 corresponds to MAX events with the top 10% MAX-day returns in the cross-section and decile 1 corresponds to events with the bottom 10% MAX returns. The sample period is from 2010 to 2022.

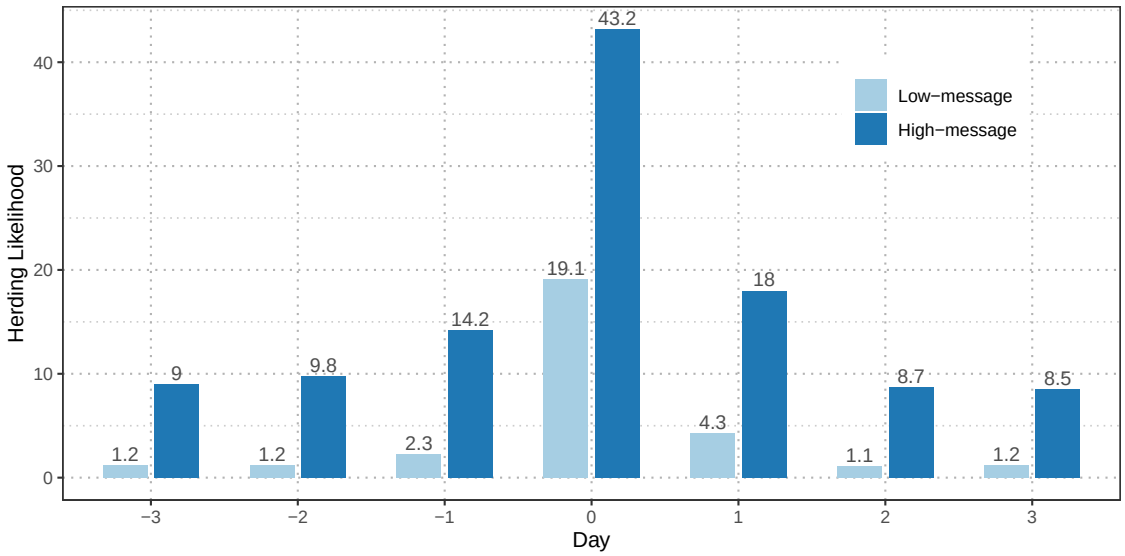


Figure 3. StockTwits Activity and Robinhood Herding around Lottery Events This figure shows daily herding likelihood around lottery events (day 0). A stock experiences a MAX event on a given day if its return on that day is the highest over a trailing 21-trading-day window. A stock is then defined as experiencing a lottery event if it has a MAX event and its MAX-day return (MAXRET) falls within the top 10% of the cross-sectional distribution. We define a herding episode as a day when the daily increase in the number of Robinhood users for a stock ranks within the top 5% of the cross-sectional distribution of the lottery stock sample, requiring a minimum threshold of 100 new users on that day. We classify each stock-day observation in the $[-3, +3]$ window around the lottery event into two groups: High-message group (prior-day messages in top 10% of the cross-sectional distribution), and the Low-message group (all others). The figure plots herding frequency for both groups. The sample period is from July 2018 to August 2020.

Table 1
Summary statistics

Panel A presents the summary statistics of each main variable utilized in the paper. Panel B displays the time series averages of the monthly cross-sectional correlations (multiplied by 100) among the variables. A stock experiences a MAX event on a given day (day 0) if its return on that day is the highest over a trailing 21-trading day window. MAXRET (in percentages) refers to the MAX-event day return. R represents the cumulative return (in percentages) over the window [1, 21]. Messages is the number of StockTwits messages over the $[-11, 0]$ window. $\mathbb{1}_{mssg}$ equals one when Message ranks in the top 10th percentile of the cross-sectional distribution, and zero otherwise. The set of stock return predictors includes the market beta (BETA), the logarithm of market capitalization measured in millions of dollars (SIZE), the logarithm of book-to-market (BM), annual asset growth rate (IA), operating profitability (OP), momentum (MOM, in percentages), past one-month return (REV, in percentages), illiquidity (ILLIQ), and co-skewness (COSKEW).

A. Descriptive Statistics

				Percentile			
Variable	Mean	Median	Stdev	10^{th}	25^{th}	75^{th}	90^{th}
R	0.80	0.29	17.83	-14.58	-6.23	6.69	15.07
MAXRET	6.42	4.41	9.32	2.00	2.86	7.24	11.96
Messages	108.17	12.00	1631.87	0.00	3.00	39.00	123.00
$\mathbb{1}_{mssg}$	0.10	0.00	0.30	0.00	0.00	0.00	0.00
BETA	1.20	1.13	0.75	0.41	0.74	1.57	2.08
SIZE	6.56	6.55	2.16	3.74	5.00	8.02	9.39
BM	-0.70	-0.58	0.96	-1.90	-1.23	-0.09	0.32
IA	0.18	0.05	1.36	-0.13	-0.03	0.16	0.41
OP	-0.01	0.02	5.10	-0.13	-0.02	0.04	0.08
MOM	12.59	6.31	68.54	-42.24	-16.85	29.82	62.75
REV	-1.31	-0.98	13.47	-15.12	-7.40	4.62	11.24
ILLIQ	3.06	0.00	65.40	0.00	0.00	0.05	0.63
COSKEW	-0.01	-0.01	0.22	-0.28	-0.15	0.13	0.26

B. Pairwise Correlations

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
(1) R	100												
(2) MAXRET	-4.3	100											
(3) Messages	0.0	10.2	100										
(4) $\mathbb{1}_{mssg}$	-1.9	18.9	15.7	100									
(5) BETA	1.1	9.5	2.1	8.5	100								
(6) SIZE	1.0	-25.1	3.7	18.6	0.6	100							
(7) BM	0.4	1.5	-3.7	-14.7	-1.5	-33.6	100						
(8) IA	-1.2	2.1	1.7	5.2	1.8	0.5	-5.4	100					
(9) OP	0.1	-0.5	0.2	0.0	0.0	0.6	1.4	-0.1	100				
(10) MOM	0.8	-5.3	8.3	3.3	6.2	5.2	2.9	-1.2	0.7	100			
(11) REV	0.0	-3.4	1.3	-1.7	-2.4	5.0	2.3	-2.1	0.5	22.8	100		
(12) ILLIQ	0.0	5.3	-0.3	-1.4	-2.6	-7.8	3.7	-0.5	-0.2	-1.8	-0.6	100	
(13) COSKEW	-1.1	0.7	1.5	5.3	16.7	3.9	-2.6	1.2	0.1	2.0	-0.3	-1.0	100

Table 2
Social interactions and lottery stock returns

This table reports the results of panel regressions of post-MAX returns on lagged explanatory variables. MAX events are defined in Table 1 and MAXRET is the MAX-day return. Cumulative returns are reported in percentages and are measured over windows [1, 21], [1, 42], and [1, 63], respectively. $\mathbb{1}_{mssg}$ equals one when the StockTwits message count for the stock over the [−11, 0] window ranks in the top 10th percentile of the cross-sectional distribution, and zero otherwise. The control variables include the market beta (BETA), the natural log of market capitalization (SIZE), the natural logarithm of book-to-market (BM), total asset growth rate (IA), operating profitability (OP), momentum (MOM), past one-month return (REV), illiquidity (ILLIQ), and co-skewness (COSKEW). The sample period is from 2010 to 2022. Firm and year-month fixed effects are included. Standard errors are two-way clustered by firm and year-month, and *t*-statistics are shown in parentheses. *p <.1; **p <.05; ***p <.01.

	Cumulative Returns		
	[1, 21]	[1, 42]	[1, 63]
	(1)	(2)	(3)
MAX $\cdot \mathbb{1}_{mssg}$	-0.146***	-0.170***	-0.248***
	(-3.55)	(-3.47)	(-3.62)
MAX	-0.052	-0.078	-0.102
	(-1.06)	(-1.33)	(-1.36)
$\mathbb{1}_{mssg}$	0.003	0.003	0.008
	(1.03)	(0.57)	(0.83)
BETA	0.001	0.002	0.001
	(0.55)	(0.57)	(0.35)
SIZE	-0.027***	-0.055***	-0.081***
	(-11.11)	(-11.60)	(-10.59)
BM	0.000	0.001	0.004
	(0.27)	(0.38)	(1.20)
IA	-0.002	-0.007**	-0.007*
	(-1.03)	(-2.06)	(-1.76)
OP	0.036***	0.057***	0.058***
	(6.23)	(5.60)	(3.65)
MOM	-0.006**	-0.018***	-0.026***
	(-2.56)	(-4.91)	(-5.24)
REV	-0.021***	-0.030**	-0.057***
	(-2.81)	(-2.45)	(-3.45)
ILLIQ	-0.012	-0.011	-0.002
	(-0.71)	(-0.48)	(-0.08)
COSKEW	-0.009	-0.017*	-0.019*
	(-1.53)	(-1.81)	(-1.82)
Obs.	430,091	426,151	421,596
Adj. R^2 (%)	10.229	15.554	17.187

Table 3
Stocktwits message volume and lottery event prediction

This table examines whether StockTwits message volume predicts the likelihood of a stock achieving lottery status and the magnitude of its lotteriness using panel regressions. MAX events are defined in Table 1 and MAXRET is the MAX-day return. A stock is then defined as experiencing a lottery event on that day if it has a MAX event and the MAX-day return (MAXRET) falls within the top 10% of the cross-sectional distribution. Columns 1 and 2 use the full sample of stock-day observations and correspond to the dependent variable, $\mathbb{1}^{lottery}$, which equals one if a stock experiences a lottery event on a given day (day 0), and zero otherwise. Columns 3 and 4 focus on a sample of stock-lottery event day observations, with the dependent variable being the lottery event day return. $\mathbb{1}_{mssg}^{preEvent}$ equals one if the StockTwits message counts mentioning the stock during the [−11,−1] day window rank in the top decile of its cross-sectional distribution. $\mathrm{R}^{preEvent}$ is the cumulative return during the [−11, −1] window. The control variables include the market beta (BETA), the natural log of market capitalization (SIZE), the natural logarithm of book-to-market (BM), total asset growth rate (IA), operating profitability (OP), momentum (MOM), past one-month return (REV), illiquidity (ILLIQ), and co-skewness (COSKEW). The sample period is from 2010 to 2022. Firm and year-month fixed effects are included. Coefficients are multiplied by 100. Standard errors are two-way clustered by firm and year-month, and t-statistics are shown in parentheses. $^{*}p<.1$; $^{**}p<.05$; $^{*}p<.01$.

	Sample: All days $\mathbb{1}^{lottery}$		Sample: Lottery-event days Lottery-event day return	
	(1)	(2)	(3)	(4)
$\mathbb{1}_{mssg}^{preEvent}$	0.228***	0.272***	3.142***	3.867***
	(10.82)	(13.33)	(5.81)	(7.09)
$\mathrm{R}^{preEvent}$	-1.033***	-1.004***	11.764***	11.258***
	(-6.81)	(-6.78)	(6.73)	(6.41)
BETA		0.045***		-0.108
		(3.67)		(-0.29)
SIZE		-0.296***		-3.119***
		(-24.47)		(-13.60)
BM		-0.012		-0.648**
		(-1.32)		(-2.58)
IA		-0.020*		-0.656**
		(-1.76)		(-2.26)
OP		-0.147***		-1.199*
		(-4.60)		(-1.68)
MOM		-0.127***		-1.958***
		(-8.36)		(-8.69)
REV		-0.460***		0.006
		(-9.00)		(0.01)
ILLIQ		0.014***		0.065**
		(9.12)		(2.17)
COSKEW		-0.007		-1.873
		(-0.26)		(-1.64)
Obs.	9,089,756	9,089,756	38,035	38,035
Adj. R^2 (%)	0.72	0.81	6.63	7.75

Table 4
Social interactions and retail herding in lottery stocks

This table presents the results of panel regressions describing the relationship between StockTwits message volume around lottery events and Robinhood herding episodes. Lottery events are defined in Table 3. Column 1's dependent variable, HR[0], indicates whether stock i experiences a herding episode on the lottery event day (day 0). The key explanatory variable, $\mathbb{1}^{preEvent}_{mssg,i}$, equals one if the stock's message volume over $[-11,-1]$ ranks in the top decile cross-sectionally, and zero otherwise. Columns 2 and 3 use HR$[1,5]$ and HR$[6,10]$ as dependent variables, representing average daily herding likelihood over days [1, 5] and [6, 10]. Their key explanatory variable, $\mathbb{1}^{0}_{mssg}$, indicates top decile message volume on day 0. Control variables are specified as follows. For column 1, Return is the cumulative return over days $[-11,-1]$, while herding indicator (HR), log number of users (USER), user changes (ΔUSER), abnormal volume (AbVol), and abnormal news (AbNews) are averaged over the same period. For columns 2 and 3, all controls are measured on day 0. EA indicates earnings announcements coinciding with lottery events. The sample period is from June 2018 to August 2020. The model includes year-month fixed effects, and standard errors are clustered by firm and year-month. t-statistics are shown in parentheses. $^{*}p<.1$; $^{**}p<.05$; $^{***}p<.01$.

	HR[0] (1)	HR$[1,5]$ (2)	HR$[6,10]$ (3)
$\mathbb{1}^{pre\text{Event}}_{mssg}$ (or $\mathbb{1}^{0}_{mssg}$)	0.070***	0.020***	0.008
	(3.34)	(2.81)	(1.47)
Return	0.100***	0.030***	0.005
	(3.78)	(4.62)	(1.05)
HR	0.606***	0.076***	0.022***
	(19.43)	(7.50)	(4.32)
USER	0.033***	0.014***	0.006***
	(7.34)	(11.00)	(6.53)
ΔUSER	0.341**	0.003	-0.006
	(2.13)	(0.42)	(-1.06)
AbVol	0.009	-0.006***	0.001
	(0.77)	(-3.20)	(0.83)
AbNews	-0.013	-0.003	-0.002*
	(-0.88)	(-1.67)	(-1.77)
EA	-0.020	-0.016***	-0.014***
	(-1.13)	(-3.26)	(-5.61)
Obs.	7,981	7,947	7,946
Adj. R^2 (%)	20.66	18.41	6.30

Table 5
Social interactions and retail trading of lottery stocks

This table presents the results of panel regressions examining the relationship between StockTwits message volume around lottery events and subsequent retail trading. Lottery events are defined in Table 3. Column 1's dependent variable, OIB[0], is the retail order imbalance on the lottery event day (day 0). The key explanatory variable, $\mathbb{1}_{mssg,i}^{preEvent}$, equals one if the stock's message volume over $[-11,-1]$ ranks in the top decile cross-sectionally, and zero otherwise. Columns 2 and 3 use OIB[1,5] and OIB[6,10] as dependent variables, representing average daily net retail order flows over days [1, 5] and [6, 10]. Their key explanatory variable, $\mathbb{1}_{mssg}^{0}$, indicates top decile message volume on day 0. Control variables are specified as follows. For column 1, Return is the cumulative return over days $[-11,-1]$, while OIB, herding indicator (HR), log number of users (USER), user changes (ΔUSER), abnormal volume (AbVol), and abnormal news (AbNews) are averaged over the same period. For columns 2 and 3, all controls are measured on day 0. EA indicates earnings announcements coinciding with lottery events. In Panel A, we use the Robinhood sample and include year-month fixed effects. In Panel B, we use the StockTwits sample and include firm and year-month fixed effects. Standard errors are clustered by firm and year-month. *t*-statistics are shown in parentheses. $^{*}p<.1$; $^{**}p<.05$; $^{***}p<.01$.

A. Robinhood Sample: 06/2018−08/2020			
	OIB[0]	OIB[1,5]	OIB[6,10]
	(1)	(2)	(3)
$\mathbb{1}_{mssg}^{preEvent}$ (or $\mathbb{1}_{mssg}^{0}$)	2.656***	1.470***	0.705
	(3.29)	(3.46)	(1.54)
Return	-3.325***	2.754***	0.800
	(-2.91)	(6.43)	(1.57)
OIB	0.039	0.043***	0.032*
	(1.16)	(4.45)	(1.78)
HR	0.988**	0.894**	1.604***
	(2.74)	(2.32)	(3.24)
USER	-0.779***	1.093***	0.804***
	(-2.94)	(9.03)	(5.51)
ΔUSER	4.260	0.427	0.365
	(1.57)	(1.12)	(1.30)
AbVol	-1.585*	-1.420***	-1.253***
	(-1.75)	(-6.78)	(-5.77)
AbNews	-3.790***	-0.292	-0.063
	(-3.30)	(-1.03)	(-0.28)
EA	1.467*	0.071	1.278
	(1.82)	(0.12)	(1.67)
Obs.	7,981	7,980	7,978
Adj. R^2 (%)	1.12	2.99	1.90

B. StockTwits Sample: 01/2010−12/2022			
	OIB[0]	OIB[1,5]	OIB[6,10]
	(1)	(2)	(3)
$\mathbb{1}_{mssg}^{preEvent}$ (or $\mathbb{1}_{mssg}^{0}$)	-0.182	1.590***	1.302***
	(-0.51)	(7.14)	(5.28)
OIB	0.042***	0.033***	0.030***
	(3.16)	(7.09)	(5.82)
Return	-2.565***	1.046***	0.009
	(-3.94)	(2.72)	(0.03)
AbVol	-0.217	-0.719***	-0.542***
	(-0.64)	(-7.39)	(-6.30)
AbNews	-0.228	-0.695***	-0.622***
	(-0.48)	(-6.52)	(-6.45)
EA	-0.766*	1.004***	2.024***
	(-1.78)	(3.25)	(5.87)
Obs.	48,327	48,094	48,041
Adj. R^2 (%)	6.73	8.13	6.71

Table 6

Social interactions and investor disagreement following lottery events

This table presents the results of panel regressions examining the relationship between StockTwits message volume around lottery events and subsequent investor disagreement. Lottery events are defined in Table 3. Column 1's dependent variable, DIS[0], is StockTwits message disagreement on the lottery event day (day 0). The key explanatory variable, $\mathbb{1}_{mssg,i}^{preEvent}$, equals one if the stock's message volume over $[-11,-1]$ ranks in the top decile cross-sectionally, and zero otherwise. Columns 2 and 3 use DIS[1,5] and DIS[6,10] as dependent variables, representing average daily StockTwits disagreement over days [1, 5] and [6, 10]. Their key explanatory variable, $\mathbb{1}_{mssg}^{0}$, indicates top decile message volume on day 0. Control variables are specified as follows. For column 1, Return is the cumulative return over days $[-11,-1]$, while DIS, abnormal volume (AbVol), and abnormal news (AbNews) are averaged over the same period. For columns 2 and 3, DIS, Return, AbVol, and AbNews are measured on day 0. EA indicates earnings announcements coinciding with lottery events. We further control for lagged market beta (BETA), log market capitalization (SIZE), log book-to-market (BM), total asset growth rate (IA), operating profitability (OP), momentum (MOM), past one-month return (REV), illiquidity (ILLIQ), and co-skewness (COSKEW). The sample period is from 2010 to 2022. We include firm and year-month fixed effect and cluster the standard errors by firm and year-month. *t*-statistics are shown in parentheses. $^{*}p<.1$; $^{**}p<.05$; $^{***}p<.01$.

	DIS[0] (1)	DIS[1,5] (2)	DIS[6,10] (3)
$\mathbb{1}_{mssg}^{preEvent}$ (or $\mathbb{1}_{mssg}^{0}$)	1.617***	1.614***	1.899***
	(6.35)	(5.80)	(6.58)
Return	0.793*	2.757***	2.315***
	(1.92)	(6.62)	(5.42)
DIS	0.113***	0.176***	0.123***
	(11.12)	(15.29)	(9.85)
AbVol	0.583***	0.782***	0.132
	(2.72)	(7.14)	(1.16)
AbNews	0.855**	0.114	-0.178
	(2.53)	(1.05)	(-1.47)
EA	-0.631**	-0.161	-0.306
	(-2.19)	(-0.44)	(-0.73)
BETA	-0.202	-0.012	-0.056
	(-1.26)	(-0.08)	(-0.32)
SIZE	0.115	0.149	0.101
	(0.75)	(1.01)	(0.54)
BM	-0.348**	-0.410***	-0.421**
	(-2.22)	(-2.68)	(-2.29)
IA	0.016	0.027	0.035
	(0.42)	(0.83)	(0.78)
OP	0.002	-0.024**	0.010
	(0.17)	(-2.34)	(0.86)
MOM	0.157***	0.265***	0.261***
	(2.71)	(4.09)	(3.13)
REV	0.306	-0.106	0.046
	(1.10)	(-0.37)	(0.17)
ILLIQ	-0.005	-0.001	0.001
	(-1.65)	(-0.31)	(0.49)
COSKEW	-1.075	-0.274	0.447
	(-1.56)	(-0.47)	(0.55)
Obs.	23,155	23,408	20,691
Adj. R^2 (%)	29.16	31.23	27.48

Table 7

Social interactions and share turnover following lottery events

This table presents the results of panel regressions examining the relationship between StockTwits message volume around lottery events and subsequent share turnover. Lottery events are defined in Table 3. Column 1's dependent variable, TO[0], is the share turnover on the lottery event day (day 0). The key explanatory variable, $\mathbb{1}_{mssg,i}^{preEvent}$, equals one if the stock's message volume over $[-11,-1]$ ranks in the top decile cross-sectionally, and zero otherwise. Columns 2 and 3 use TO[1,5] and TO[6,10] as dependent variables, representing average daily share turnover over days [1, 5] and [6, 10]. Their key explanatory variable, $\mathbb{1}_{mssg}^{0}$, indicates top decile message volume on day 0. Control variables are specified as follows. For column 1, Return is the cumulative return over days $[-11,-1]$, while TO, abnormal volume (AbVol), and abnormal news (AbNews) are averaged over the same period. For columns 2 and 3, TO, Return, AbVol, and AbNews are measured on day 0. EA indicates earnings announcements coinciding with lottery events. We further control for lagged market beta (BETA), log market capitalization (SIZE), log book-to-market (BM), total asset growth rate (IA), operating profitability (OP), momentum (MOM), past one-month return (REV), illiquidity (ILLIQ), and co-skewness (COSKEW). The dependent variables are in percentage points. The sample period is from 2010 to 2022. We include firm and year-month fixed effect and cluster the standard errors by firm and year-month. *t*-statistics are shown in parentheses. *p <.1; **p <.05; ***p <.01.

	TO[0] (1)	TO[1,5] (2)	TO[6,10] (3)
$\mathbb{1}_{mssg}^{preEvent}$ (or $\mathbb{1}_{mssg}^{0}$)	10.681***	0.890***	0.721***
	(6.20)	(15.32)	(13.04)
Return	9.593***	1.163***	0.961***
	(3.91)	(10.30)	(10.33)
TO	1.781***	0.384***	0.260***
	(5.07)	(50.25)	(32.26)
AbVol	2.913***	-0.064***	-0.124***
	(3.04)	(-5.76)	(-10.63)
AbNews	0.714	-0.058***	-0.078***
	(0.83)	(-5.99)	(-8.22)
EA	-1.935*	-0.217***	-0.148***
	(-1.71)	(-7.76)	(-5.57)
BETA	0.698	0.032	0.036
	(0.82)	(1.18)	(1.23)
SIZE	-7.925***	-0.122***	-0.140***
	(-9.13)	(-5.66)	(-5.77)
BM	0.135	-0.037*	-0.032
	(0.17)	(-1.68)	(-1.36)
IA	-0.072	0.010*	0.013**
	(-0.43)	(1.92)	(2.59)
OP	-0.036	0.001	0.000
	(-1.26)	(0.63)	(0.35)
MOM	-1.970***	0.083***	0.099***
	(-4.61)	(6.90)	(6.87)
REV	4.059**	0.111**	0.098*
	(2.21)	(2.23)	(1.68)
ILLIQ	-0.002**	-0.001	-0.001
	(-2.29)	(-1.04)	(-0.80)
COSKEW	-7.256**	0.003	0.054
	(-2.25)	(0.04)	(0.61)
Obs.	35,107	35,084	35,024
Adj. R^2 (%)	26.31	58.17	50.57

Table 8
Headquarters county social connectedness and lottery stock returns

This table reports the results of panel regressions of post-MAX returns on lagged explanatory variables. MAX events are defined in Table 1 and MAXRET is the MAX-day return. Cumulative returns are reported in percentages and are measured over windows [1,21], [1,42], and [1,63], respectively. SCIH is the social connectedness of the stock's firm headquarters county. The control variables include the market beta (BETA), the natural log of market capitalization (SIZE), the natural logarithm of book-to-market (BM), total asset growth rate (IA), operating profitability (OP), momentum (MOM), past one-month return (REV), illiquidity (ILLIQ, multiplied by 10^{-2}), and co-skewness (COSKEW). The sample period is from 2010 to 2022. Firm and year-month fixed effects are included. Standard errors are two-way clustered by firm and year-month, and t-statistics are shown in parentheses. *p <.1; **p <.05; ***p <.01.

	Cumulative Return		
	[1, 21]	[1, 42]	[1, 63]
	(1)	(2)	(3)
	(1)	(2)	(3)
MAXRET · SCIH	-0.023**	-0.031**	-0.033*
	(-2.31)	(-2.09)	(-1.90)
MAXRET	-0.079	-0.099	-0.128*
	(-1.56)	(-1.53)	(-1.70)
SCIH	0.005***	0.007**	0.008**
	(3.74)	(2.60)	(2.01)
BETA	0.000	0.000	0.000
	(0.19)	(0.05)	(-0.04)
SIZE	-0.027***	-0.054***	-0.080***
	(-10.94)	(-12.33)	(-12.09)
BM	0.000	0.000	0.002
	(0.08)	(-0.18)	(0.60)
IA	0.000	-0.001	-0.003*
	(-0.04)	(-1.24)	(-1.70)
OP	0.003***	0.004**	0.003*
	(3.52)	(2.56)	(1.87)
MOM	-0.006***	-0.014***	-0.021***
	(-3.07)	(-4.51)	(-5.16)
REV	-0.019**	-0.034***	-0.061***
	(-2.53)	(-2.83)	(-3.76)
ILLIQ	-0.008	-0.008	0.008
	(-0.30)	(-0.25)	(0.21)
COSKEW	-0.009	-0.015*	-0.017*
	(-1.53)	(-1.73)	(-1.79)
Obs.	424,628	423,609	422,444
Adj. R^2 (%)	11.50	16.89	18.72

Table 9
Lottery event prediction: 2SLS analysis

This table presents two-stage least squares (2SLS) estimation results using state-wide social media privacy law implementation as an instrument for social media discussion. The sample includes all stock-day observations from 2010 to 2022. Columns 1 and 3 report first-stage results where the dependent variable is $\mathbb{1}_{mssg}^{preEvent}$, an indicator equal to 1 if the number of StockTwits messages during the $[-11, -1]$ window belongs to the top 10 percent of the distribution. Columns 2 and 4 report second-stage results with lottery event dummy and MAX-day return as dependent variables, respectively. SMP equals 1 for firms headquartered in states that have implemented social media privacy laws protecting employee social media use from employer monitoring. Control variables include lagged values of pre-event cumulative return ($R^{preEvent}$), market beta (BETA), log firm size (SIZE), log book-to-market ratio (BM), asset growth (IA), operating profitability (OP), momentum (MOM), short-term reversal (REV), illiquidity (ILLIQ), and coskewness (COSKE). Firm and year-month fixed effects are included. Coefficients are multiplied by 100. Standard errors are two-way clustered by firm and year-month, and t-statistics are shown in parentheses. $^{*}p < .1$; $^{**}p < .05$; $^{*}p < .01$.

	Sample: All days		Sample: Lottery event days	
	Dependent variable		Dependent variable	
	(1)	(2)	(3)	(4)
	First Stage	Second Stage	First Stage	Second Stage
	$\mathbb{1}_{mssg}^{preEvent}$	$\mathbb{1}^{lottery}$	$\mathbb{1}_{mssg}^{preEvent}$	Lottery-day return
$\mathbb{1}_{mssg}^{preEvent}$		2.790**		12.568
		(2.54)		(0.90)
SMP	0.014***		0.029**	
	(2.87)		(2.59)	
$R^{preEvent}$	0.153***	−1.406***	0.027*	9.854***
	(15.24)	(−6.20)	(1.70)	(6.49)
BETA	0.022***	−0.006	0.031***	−0.221
	(5.42)	(−0.19)	(4.38)	(−0.46)
SIZE	0.004	−0.305***	0.010*	−3.195***
	(1.24)	(−17.91)	(1.81)	(−11.99)
BM	−0.011***	0.012	−0.021***	−0.412
	(−2.68)	(0.82)	(−3.53)	(−1.13)
IA	0.016***	−0.064***	0.020***	−0.847*
	(5.98)	(−3.03)	(2.97)	(−1.74)
OP	−0.051***	−0.023	−0.032*	−0.839
	(−5.64)	(−0.39)	(−1.98)	(−0.97)
MOM	0.030***	−0.203***	0.040***	−2.361***
	(6.46)	(−5.58)	(6.50)	(−3.58)
REV	0.046***	−0.588***	0.029**	−0.456
	(11.07)	(−7.95)	(2.50)	(−0.50)
ILLIQ	−0.001***	0.016***	−0.001**	0.077**
	(−4.72)	(8.21)	(−2.35)	(2.11)
COSKEW	−0.012*	0.033	−0.025	−1.593**
	(−1.73)	(1.15)	(−1.45)	(−2.49)
F-stat	1,839.7		12.7	
Obs.	8,822,343	8,822,343	35,472	35,472
Adj. R^2 (%)	46.03	0.02	41.56	6.73

Table 10
Social interactions and lottery stocks: Controlling for attention

This table examines the relationship between StockTwits message activity and lottery stocks while controlling for alternative attention measures. Panel A examines post-MAX event returns over 21, 42, and 63 trading days. Panel B examines lottery event prediction. ASV is the abnormal Google search volume and DMR is the Bloomberg readership. For Panel A, both ASV and DMR are averaged over the $[-11, 0]$ window. For Panel B, both are measured over the pre-event window $[-11, -1]$. MAX is the return on the MAX event day. $\mathbb{1}_{mssg}$ equals one if the StockTwits message counts rank in the top decile of the cross-sectional distribution during the $[-11, 0]$ window. A stock experiences a lottery event if it has a MAX event and the MAX-day return (MAXRET) falls within the top 10% of the cross-sectional distribution. The sample period is from 2010 to 2022. All specifications include firm and year-month fixed effects and the full set of control variables. Coefficients in Panel B are multiplied by 100. Standard errors are two-way clustered by firm and year-month, and t-statistics are shown in parentheses. $^{*}p < .1$; $^{**}p < .05$; $^{***}p < .01$.

A. Post-MAX Event Returns

	Cumulative Returns		
	[1, 21]	[1, 42]	[1, 63]
	(1)	(2)	(3)
MAX·$\mathbb{1}_{mssg}$	-0.162***	-0.182***	-0.263***
	(-3.64)	(-3.49)	(-3.73)
MAX·ASV	0.059	-0.166	-0.212
	(0.48)	(-1.10)	(-1.14)
MAX·DMR	0.080**	0.114**	0.141**
	(2.45)	(2.50)	(2.43)
MAX	-0.068	-0.096*	-0.124*
	(-1.47)	(-1.78)	(-1.81)
$\mathbb{1}_{mssg}$	0.004	0.003	0.009
	(1.12)	(0.65)	(0.87)
ASV	0.002	0.009	0.013
	(0.27)	(1.24)	(1.42)
DMR	-0.003	-0.005*	-0.006*
	(-1.60)	(-1.87)	(-1.81)
BETA	0.001	0.002	0.001
	(0.55)	(0.59)	(0.36)
SIZE	-0.027***	-0.055***	-0.082***
	(-11.21)	(-11.66)	(-10.62)
BM	0.000	0.001	0.004
	(0.29)	(0.39)	(1.21)
IA	-0.002	-0.007**	-0.007*
	(-1.06)	(-2.07)	(-1.77)
OP	0.036***	0.057***	0.059***
	(6.24)	(5.70)	(3.75)
MOM	-0.006**	-0.018***	-0.026***
	(-2.56)	(-4.95)	(-5.25)
REV	-0.022***	-0.030**	-0.056***
	(-3.00)	(-2.47)	(-3.43)
ILLIQ	-0.011	-0.010	-0.001
	(-0.66)	(-0.45)	(-0.04)
COSKE	-0.009	-0.017*	-0.019*
	(-1.55)	(-1.83)	(-1.82)
Obs.	429,002	425,079	420,526
Adj. R^2 (%)	10.23	15.57	17.20

B. Lottery Event Prediction		
	Sample: All days $\mathbb{1}^{lottery}$	Sample: Lottery event days Lottery-event day return
	(1)	(2)
$\mathbb{1}_{mssg}^{preEvent}$	0.247***	3.598***
	(12.07)	(6.34)
$\text{ASV}^{preEvent}$	0.064**	3.581***
	(2.10)	(2.70)
$\text{DMR}^{preEvent}$	0.089***	1.010***
	(11.35)	(3.14)
$\text{R}^{preEvent}$	-1.004***	11.264***
	(-6.79)	(6.45)
BETA	0.045***	-0.112
	(3.64)	(-0.30)
SIZE	-0.301***	-3.167***
	(-24.71)	(-13.75)
BM	-0.012	-0.644**
	(-1.36)	(-2.57)
IA	-0.021*	-0.657**
	(-1.82)	(-2.26)
OP	-0.146***	-1.188*
	(-4.56)	(-1.66)
MOM	-0.127***	-1.983***
	(-8.39)	(-8.78)
REV	-0.461***	-0.051
	(-9.03)	(-0.06)
ILLIQ	0.014***	0.065**
	(9.10)	(2.18)
COSKE	-0.005	-1.858
	(-0.19)	(-1.63)
Obs.	9,089,756	38,035
Adj. R^2 (%)	0.81	7.78

Appendix

- Table A1 presents monthly tests of the relation of MAXRET with future stock returns using univariate portfolio sorts and Fama-MacBeth regressions.
- Table A2 presents panel regression tests of the relation of MAXRET with future stock returns.
- Table A3 repeats the tests in Table 2 using alternative StockTwits data shared by Cookson and Niessner (2023), covering 2010–2021.
- Table A4 presents the test of social interactions and lottery stock returns around future earnings announcements.
- Table A5 repeats the tests in Table 2 using an alternative lottery measure adapted from Kumar (2009).
- Table A6 repeats the tests in Table 3 using an alternative lottery measure adapted from Kumar (2009).
- Table A7 reproduces Table 2 using alternative message windows.
- Table A8 extends Table 2 by controlling investor attention, information supply, and arbitrage costs.

Table A1
Monthly analysis of maxret (max) and future stock returns

In Panel A, we report the average monthly returns of portfolios sorted on MAXRET. MAX events are defined in Table 1 and MAXRET is the MAX-day return. If a stock experiences multiple MAX events for a given month, we set MAXRET to the highest MAXRET for the month. Columns 1–10 present the value-weighted averages of the following variables (in percentage points): MAXRET, the excess return (RET−RF), the Fama-French-Carhart-Pastor-Stambaugh alpha (FFCPS), and the Fama-French five-factor alphas (FF5). The column "High−Low" is the difference between Decile 10 (High) and Decile 1 (Low) in the corresponding variables. The excess market returns and the size, book-to-market, momentum, profitability, and investment factors are from Kenneth French's data library. The liquidity factor is from Lubos Pastor's data library. Panel B presents results of the Fama-MacBeth regression, where one-, two- and three-month-ahead stock returns are regressed on MAXRET and a set of control variables, including the market beta (BETA), the natural log of market capitalization (SIZE), the natural logarithm of book-to-market (BM), total asset growth rate (IA), operating profitability (OP), momentum (MOM), past one-month return (REV), illiquidity (ILLIQ), and co-skewness (COSKEW). Panel C repeats the Fama-MacBeth regression for one-, two-, and three-month-ahead stock returns, using the highest daily return in a month (MAX) to capture a stock's lottery-like characteristics, following Bali, Cakici, and Whitelaw (2011). The Newey-West adjusted t-statistics are given in parentheses. The sample period for the univariate portfolio analysis spans July 1963 through December 2022. The Fama-MacBeth regression analysis covers the period from February 1972 to December 2022, depending on the availability of the quarterly earnings announcement date (quarterly Compustat item "RDQ"). As a result, the operating profitability measure (OP) begins in 1972. $^{*}p<.1$; $^{**}p<.05$; $^{***}p<.01$.

A. Univariate Portfolio Analysis

	1 (Low)	2	3	4	5	6	7	8	9	10 (High)	High−Low
MAXRET	1.63	2.75	3.63	4.50	5.44	6.55	7.95	9.88	13.11	24.03	22.40
RET−RF	0.52***	0.57***	0.50***	0.60***	0.70***	0.62***	0.56**	0.30	0.09	-0.46	-0.99***
	(3.69)	(3.65)	(2.89)	(3.19)	(3.16)	(2.64)	(2.02)	(0.97)	(0.28)	(-1.36)	(-3.38)
FFCPS	0.06	0.09	-0.04	0.01	0.03	0.04	-0.07	-0.36***	-0.60***	-1.14***	-1.20***
	(0.72)	(1.50)	(-0.68)	(0.18)	(0.32)	(0.43)	(-0.67)	(-3.17)	(-3.93)	(-5.37)	(-5.02)
FF5	-0.04	0.00	-0.11**	0.02	0.10	0.13*	0.03	-0.17*	-0.33***	-0.86***	-0.82***
	(-0.48)	(0.00)	(-1.96)	(0.25)	(1.27)	(1.71)	(0.32)	(-1.74)	(-2.84)	(-5.21)	(-4.52)

B. Fama-MacBeth Regressions Using MAXRET

	Cumulative Excess Returns		
	One-month	Two-month	Three-month
	(1)	(2)	(3)
MAXRET	-0.027***	-0.064***	-0.091***
	(-4.15)	(-5.31)	(-4.90)
BETA	0.118	0.197	0.295
	(1.23)	(1.05)	(1.06)
SIZE	-0.071**	-0.147**	-0.231**
	(-2.07)	(-2.28)	(-2.42)
BM	0.206***	0.445***	0.704***
	(3.10)	(3.57)	(3.76)
IA	-0.439***	-0.800***	-1.156***
	(-7.80)	(-7.33)	(-7.67)
OP	1.361***	2.572***	3.344***
	(4.67)	(5.39)	(5.48)
MOM	0.006***	0.011***	0.016***
	(4.57)	(4.50)	(4.44)
REV	-0.047***	-0.041***	-0.028***
	(-8.81)	(-5.69)	(-3.22)
ILLIQ	0.008***	0.014***	0.019***
	(3.92)	(3.65)	(3.63)
COSKEW	-0.131	-0.267	-0.406
	(-1.24)	(-1.38)	(-1.46)
Obs.	1,387,880	1,378,039	1,367,606
Adj. R^2 (%)	5.58	6.14	6.38

C. Fama-MacBeth Regressions Using MAX

	Cumulative Excess Returns		
	One-month	Two-month	Three-month
	(1)	(2)	(3)
MAX	-0.031***	-0.071***	-0.102***
	(-4.61)	(-5.52)	(-5.15)
BETA	0.106	0.214	0.311
	(1.12)	(1.15)	(1.14)
SIZE	-0.070**	-0.150**	-0.231**
	(-2.11)	(-2.33)	(-2.42)
BM	0.208***	0.439***	0.685***
	(3.19)	(3.47)	(3.67)
IA	-0.434***	-0.798***	-1.162***
	(-7.78)	(-7.52)	(-7.84)
OP	1.181***	2.154***	2.792***
	(4.23)	(4.59)	(4.66)
MOM	0.006***	0.011***	0.016***
	(4.51)	(4.62)	(4.42)
REV	-0.047***	-0.040***	-0.025***
	(-8.95)	(-5.53)	(-2.79)
ILLIQ	0.008***	0.016***	0.020***
	(4.32)	(4.00)	(4.08)
COSKEW	-0.149	-0.245	-0.353
	(-1.41)	(-1.27)	(-1.29)
Obs.	1,971,676	1,957,317	1,942,969
Adj. R^2 (%)	5.40	5.97	6.21

Table A2
Panel regression analysis of maxret and future futures: Baseline model

This table reports the results of panel regressions of post-MAX returns on lagged explanatory variables. MAX events are defined in Table 1 and MAXRET is the MAX-day return. Cumulative returns are reported in percentages and are measured over windows [1, 21], [1, 42], and [1, 63], respectively. The control variables include the market beta (BETA), the natural log of market capitalization (SIZE), the natural logarithm of book-to-market (BM), total asset growth rate (IA), operating profitability (OP), momentum (MOM), past one-month return (REV), illiquidity (ILLIQ, multiplied by 10^{-2}), and co-skewness (COSKEW). The sample period is from 2010 to 2022. Firm and year-month fixed effects are included. Standard errors are two-way clustered by firm and year-month, and t-statistics are shown in parentheses. $^{*}p<.1$; $^{**}p<.05$; $^{***}p<.01$.

	Cumulative Return		
	[1, 21]	[1, 42]	[1, 63]
	(1)	(2)	(3)
MAXRET	-0.104**	-0.141**	-0.187***
	(-2.32)	(-2.56)	(-2.68)
BETA	0.001	0.002	0.001
	(0.48)	(0.51)	(0.30)
SIZE	-0.027***	-0.055***	-0.081***
	(-11.18)	(-11.72)	(-10.74)
BM	0.001	0.001	0.004
	(0.40)	(0.48)	(1.29)
IA	-0.002	-0.007**	-0.008*
	(-1.07)	(-2.10)	(-1.80)
OP	0.037***	0.058***	0.059***
	(6.36)	(5.68)	(3.70)
MOM	-0.006***	-0.018***	-0.026***
	(-2.72)	(-5.03)	(-5.29)
REV	-0.021***	-0.031**	-0.058***
	(-2.87)	(-2.49)	(-3.49)
ILLIQ	-0.006	-0.005	0.006
	(-0.38)	(-0.21)	(0.21)
COSKEW	-0.009	-0.017*	-0.019*
	(-1.52)	(-1.81)	(-1.83)
Obs.	430,091	426,151	421,596
Adj. R^2 (%)	10.18	15.52	17.15

Table A3
Social interactions and lottery stock returns: Alternative stocktwits sample

This table repeats Table 2 using the alternative StockTwits data shared by Cookson and Niessner (2023) for the period from 2010 to 2021. See Table 2 for regression specification and variable descriptions.

	Cumulative Return		
	[1, 21]	[1, 42]	[1, 63]
	(1)	(2)	(3)
MAXRET $\cdot \mathbb{1}_{mssg}$	-0.122**	-0.181***	-0.257***
	(-2.04)	(-3.35)	(-3.63)
MAXRET	-0.065	-0.077	-0.104
	(-1.16)	(-1.17)	(-1.26)
$\mathbb{1}_{mssg}$	0.005	0.009**	0.015
	(1.09)	(2.02)	(1.44)
BETA	0.001	0.003	0.002
	(0.60)	(0.76)	(0.51)
SIZE	-0.028***	-0.058***	-0.086***
	(-10.62)	(-12.02)	(-11.25)
BM	0.001	0.002	0.005
	(0.70)	(0.75)	(1.47)
IA	0.000	-0.004	-0.005
	(0.03)	(-1.46)	(-1.28)
OP	0.035***	0.053***	0.058***
	(5.82)	(5.20)	(3.69)
MOM	-0.009***	-0.023***	-0.033***
	(-3.73)	(-6.54)	(-6.94)
REV	-0.016**	-0.025*	-0.050***
	(-2.18)	(-1.95)	(-2.97)
ILLIQ	-0.022	-0.022	-0.014
	(-1.38)	(-0.95)	(-0.48)
COSKEW	-0.010	-0.018*	-0.017
	(-1.63)	(-1.88)	(-1.62)
Obs.	384,745	383,852	382,791
Adj. R^2 (%)	10.25	15.79	17.70

Table A4
Social interactions and lottery stock returns around future earnings announcements

This table presents results of a panel regression examining the relationship between StockTwits message activity surrounding MAX event days and subsequent earnings announcement returns. MAX events are defined in Table 1 and MAXRET is the MAX-day return. The dependent variable is the DGTW-adjusted cumulative abnormal returns (in percentages), for the three days surrounding the subsequent quarter's earnings announcement. $\mathbb{1}_{mssg}$ equals one if the stock ranks in the top 10% for messages posted during the $[-11, 0]$ window. The control variables include the market beta (BETA), the natural log of market capitalization (SIZE), the natural logarithm of book-to-market (BM), total asset growth rate (IA), operating profitability (OP), momentum (MOM), past one-month return (REV), illiquidity (ILLIQ, multiplied by 10^{-2}), and co-skewness (COSKEW). The sample period is from 2010 to 2022. Firm and year-month fixed effects are included. Standard errors are two-way clustered by firm and year-month, and the resultant t-statistics are shown in parentheses. *p <.1; **p <.05; ***p <.01.

	CAR$[-1,1]$	
	(1)	(2)
MAXRET · $\mathbb{1}_{mssg}$		-2.517**
		(-2.53)
MAXRET	-3.010***	-2.069***
	(-5.02)	(-3.00)
$\mathbb{1}_{mssg}$		0.175
		(1.19)
BETA	0.069	0.071
	(0.79)	(0.81)
SIZE	-1.135***	-1.135***
	(-15.00)	(-15.03)
BM	-0.030	-0.033
	(-0.48)	(-0.52)
IA	0.026	0.026
	(0.95)	(0.93)
OP	-0.002	-0.002
	(-0.36)	(-0.36)
MOM	-0.257***	-0.252***
	(-4.45)	(-4.39)
REV	-0.531**	-0.514*
	(-2.04)	(-1.97)
ILLIQ	-0.019	-0.020
	(-0.59)	(-0.64)
COSKEW	-0.133	-0.142
	(-0.72)	(-0.77)
Obs.	128,088	127,992
Adj. R^2 (%)	6.4	6.4

Table A5

Social interactions and lottery stock returns: Alternative lottery measure

This table examines the relationship between StockTwits message activity and post-lottery event returns using a lottery index adapted from Kumar (2009). For each month, we sort stocks into 50 bins by price per share (PRC) in descending order, such that stocks in the lowest bin (i.e., a PRC portfolio rank of 1) have the highest price per share and those in the highest bin (i.e., a PRC portfolio rank of 50) have the lowest price per share. We also independently sort stocks into 50 bins by idiosyncratic volatility (IVOL) and idiosyncratic skewness (ISKEW) in ascending order. IVOL and ISKEW are, respectively, the standard deviation and the skewness of residuals from the time series regression of daily stock returns against the daily market, size, and book-to-market factors in a month. The three ranks are summed to create a composite lottery index (SumRank). We then define LTRY as decile rank of SumRank, with the highest decile (LTRY = 10) representing the most lottery-like stocks. The dependent variables are cumulative returns over 21, 42, and 63 trading days. $\mathbb{1}_{mssg}$ equals one if the StockTwits message counts during the $[-11, 0]$ window rank in the top decile of the cross-sectional distribution. Control variables include lagged monthly return (REV), market beta (BETA), log firm size (SIZE), log book-to-market ratio (BM), asset growth (IA), operating profitability (OP), momentum (MOM), illiquidity (ILLIQ), and co-skewness (COSKE). The sample period is from 2010 to 2022. Firm and year-month fixed effects are included. Standard errors are two-way clustered by firm and year-month, and t-statistics are shown in parentheses. $^{*}p<.1$; $^{**}p<.05$; $^{***}p<.01$.

	Cumulative Returns		
	[1,21]	[1,42]	[1,63]
	(1)	(2)	(3)
LTRY· $\mathbb{1}_{mssg}$	-0.127**	-0.271***	-0.365***
	(-2.21)	(-3.09)	(-2.97)
LTRY	-0.009	-0.050	-0.069
	(-0.41)	(-1.42)	(-1.54)
$\mathbb{1}_{mssg}$	0.206	0.501	0.579
	(0.76)	(1.17)	(1.04)
BETA	0.091	0.180	0.290
	(0.47)	(0.60)	(0.79)
SIZE	-2.762***	-5.479***	-7.998***
	(-12.35)	(-13.20)	(-13.00)
BM	0.010	0.102	0.198
	(0.08)	(0.52)	(0.73)
IA	-0.285*	-0.594**	-1.044***
	(-1.76)	(-2.42)	(-2.97)
OP	2.693***	3.532***	3.633***
	(6.14)	(4.49)	(3.27)
MOM	-0.828***	-1.645***	-2.517***
	(-3.32)	(-4.67)	(-5.59)
REV	-1.236**	-2.424***	-3.405***
	(-2.03)	(-2.73)	(-3.58)
ILLIQ	0.829*	1.843*	1.774
	(1.73)	(1.80)	(1.52)
COSKEW	-0.415	-1.151	-1.617
	(-0.68)	(-1.28)	(-1.64)
Obs.	434,716	433,700	432,570
Adj. R^2 (%)	13.71	15.24	16.87

Table A6
Stocktwits message volume and lottery event prediction: Alternative lottery measure

This table examines whether StockTwits message volume predicts the likelihood of a stock achieving lottery status using a lottery index adapted from Kumar (2009). For each month, we sort stocks into 50 bins by price per share (PRC) in descending order, such that stocks in the lowest bin (i.e., a PRC portfolio rank of 1) have the highest price per share and those in the highest bin (i.e., a PRC portfolio rank of 50) have the lowest price per share. We also independently sort stocks into 50 bins by idiosyncratic volatility (IVOL) and idiosyncratic skewness (ISKEW) in ascending order. IVOL and ISKEW are, respectively, the standard deviation and the skewness of residuals from the time series regression of daily stock returns against the daily market, size, and book-to-market factors in a month. The three ranks are summed to create a composite lottery index (SumRank). A lottery event occurs when a stock's SumRank ranking is in the top decile. Columns 1 and 2 use the full sample of stock-day observations and correspond to the dependent variable, $\mathbb{1}^{lottery}$, which equals one if a stock experiences a lottery event, and zero otherwise. $\mathbb{1}^{preMonth}_{mssg}$ equals one if the StockTwits message counts mentioning the stock during the previous month rank in the top decile of its cross-sectional distribution. $\mathrm{R}^{preEvent}$ is the cumulative return during the $[-11, -1]$ window. The control variables include the market beta (BETA), the natural log of market capitalization (SIZE), the natural logarithm of book-to-market (BM), total asset growth rate (IA), operating profitability (OP), momentum (MOM), past one-month return (REV), illiquidity (ILLIQ), and co-skewness (COSKE). The sample period is from 2010 to 2022. Firm and year-month fixed effects are included. Coefficients are multiplied by 100. Standard errors are two-way clustered by firm and year-month, and t-statistics are shown in parentheses. $^{*}p<.1$; $^{**}p<.05$; $^{***}p<.01$.

	Sample: All days $\mathbb{1}^{lottery}$	
	(1)	(2)
$\mathbb{1}^{preMonth}_{mssg}$	0.022***	0.029***
	(7.13)	(10.94)
$\mathrm{R}^{preEvent}$	-0.056***	-0.009
	(-6.11)	(-1.23)
BETA		0.007***
		(3.43)
SIZE		-0.067***
		(-30.83)
BM		0.002
		(1.02)
IA		-0.011***
		(-5.25)
OP		-0.022***
		(-4.40)
MOM		-0.046***
		(-17.69)
REV		-0.029***
		(-4.03)
ILLIQ		0.049***
		(6.44)
COSKEW		0.001
		(0.21)
Obs.	501,340	435,690
Adj. R^2 (%)	25.36	27.86

Table A7
Social interactions and lottery stock returns: Alternative message windows

This table reproduces Table 2 using alternative definitions of StockTwits message indicator, $\mathbb{1}_{mssg}$, that are measured with different event windows: $[-5, 0]$, $[-3, 0]$, and $[-1, 0]$, respectively. See Table 2 for regression specification and variable definitions.

	Cumulative Return		
	[1, 21]	[1, 42]	[1, 63]
	(1)	(2)	(3)
A. Message Window: [-5, 0]			
MAXRET · $\mathbb{1}_{mssg}$	-0.103*	-0.155***	-0.192***
	(-1.93)	(-3.45)	(-2.90)
MAXRET	-0.056	-0.069	-0.105
	(-1.05)	(-1.12)	(-1.35)
$\mathbb{1}_{mssg}$	0.003	0.004	0.008
	(0.72)	(0.98)	(0.96)
Obs.	429,002	425,079	420,526
Adj. R^2 (%)	10.20	15.55	17.17
B. Message Window: [-3, 0]			
MAXRET · $\mathbb{1}_{mssg}$	-0.101*	-0.141***	-0.155**
	(-1.90)	(-2.86)	(-2.27)
MAXRET	-0.048	-0.063	-0.107
	(-0.82)	(-0.94)	(-1.24)
$\mathbb{1}_{mssg}$	0.002	0.002	0.005
	(0.49)	(0.59)	(0.64)
Obs.	429,002	425,079	420,526
Adj. R^2 (%)	10.20	15.55	17.17
C. Message Window: [-1, 0]			
MAXRET · $\mathbb{1}_{mssg}$	-0.083	-0.136***	-0.136*
	(-1.47)	(-2.63)	(-1.90)
MAXRET	-0.055	-0.056	-0.103
	(-0.81)	(-0.75)	(-1.04)
$\mathbb{1}_{mssg}$	0.004	0.004	0.004
	(1.26)	(1.37)	(0.74)
Obs.	429,002	425,079	420,526
Adj. R^2 (%)	10.19	15.54	17.16

Table A8
Social interactions and lottery stock returns: Additional controls

This table extends Table 2 by controlling for analyst coverage (CVRG) in Panel A, firm-level news (NEWS) in Panel B , and arbitrage costs measured by idiosyncratic volatility (IVOL) in Panel C, and arbitrage cost index (COST) in Panel D, respectively. See Table 2 for regression specification and variable definitions.

	Cumulative Return		
	[1, 21]	[1, 42]	[1, 63]
	(1)	(2)	(3)
A. Investor Attention: CVRG			
MAXRET $\cdot \mathbb{1}_{mssg}$	-0.144***	-0.167***	-0.245***
	(-3.48)	(-3.42)	(-3.61)
CVRG	0.000	0.000	-0.002
	(-0.02)	(-0.07)	(-0.40)
Obs.	428,979	425,056	420,503
Adj. R^2 (%)	10.22	15.58	17.19
B. Information Supply: NEWS			
MAXRET $\cdot \mathbb{1}_{mssg}$	-0.156***	-0.190***	-0.285***
	(-3.27)	(-3.68)	(-3.70)
NEWS	-0.001	-0.003***	-0.005***
	(-1.33)	(-2.75)	(-3.13)
Obs.	357,905	357,047	356,058
Adj. R^2 (%)	10.39	15.32	17.59
C. Arbitrage Cost: IVOL			
MAXRET $\cdot \mathbb{1}_{mssg}$	-0.143***	-0.166***	-0.244***
	(-3.46)	(-3.39)	(-3.61)
IVOL	0.000	-0.001	-0.001
	(-0.94)	(-1.13)	(-0.65)
Obs.	428,979	425,056	420,503
Adj. R^2 (%)	10.22	15.59	17.19
D. Arbitrage Cost: COST			
MAXRET $\cdot \mathbb{1}_{mssg}$	-0.139***	-0.158***	-0.232***
	(-3.36)	(-3.23)	(-3.42)
COST	0.004***	0.008***	0.011***
	(2.71)	(3.13)	(3.63)
Obs.	428,979	425,056	420,503
Adj. R^2 (%)	10.23	15.61	17.22

Social Interaction and Lottery Stock Mania

Internet Appendix

- Table S.I lists lottery stocks whose market capitalizations fall within the top NYSE size decile.
- Table S.II reports the value-weighted returns of MAXRET-based decile portfolios formed from stocks with market capitalizations above the NYSE median size breakpoint.
- Table S.III presents the implementation dates and statutory citations of state-level media privacy laws.
- Table S.IV examines the robustness of Table 2 to alternative definitions of lottery events.
- Table S.V examines the robustness of Table 3 to alternative definitions of lottery events.
- Table S.VI examines the robustness of Table 4 to alternative definitions of lottery events.
- Table S.VII examines the robustness of Table 5 to alternative definitions of lottery events.

Table S.I
Lottery stocks in the top NYSE size decile

Each portfolio formation month, stocks are sorted by MAXRET, as defined in Table 1. If a stock experiences multiple MAX events within a month, its MAXRET is assigned the highest value observed that month. Lottery stocks are defined as those in the top decile of the monthly MAXRET distribution. This table lists lottery stocks whose market capitalizations fall within the top NYSE size decile in each portfolio formation month from 2010 to 2022. A total of 148 such stock-month observations are identified during this period. The last two columns report the average daily abnormal Google search volume (ASV) over the window from day −11 to day 0 (the MAXRET day), and an indicator for abnormal StockTwits activity ($\mathbb{1}_{mssg}$) over the same period, which equals 1 if the number of StockTwits messages during the window ranks in the top 10% cross-sectionally, and 0 otherwise.

DATE	Company Name	Ticker Symbol	ASV	$\mathbb{1}_{mssg}$
1/31/2018	ABBVIE INC	ABBV	57.31%	1
12/31/2013	ADOBE SYSTEMS INC	ADBE	21.76%	1
8/31/2018	AUTODESK INC	ADSK	N/A	1
4/30/2014	ALLERGAN INC	AGN	48.26%	0
10/30/2020	ALIGN TECHNOLOGY INC	ALGN	N/A	1
7/31/2013	ALEXION PHARMACEUTICALS INC	ALXN	N/A	1
1/31/2014	ALEXION PHARMACEUTICALS INC	ALXN	N/A	1
12/31/2020	ALEXION PHARMACEUTICALS INC	ALXN	N/A	0
5/31/2016	APPLIED MATERIALS INC	AMAT	20.86%	1
4/30/2012	AMAZON COM INC	AMZN	8.24%	1
1/30/2015	AMAZON COM INC	AMZN	32.96%	1
4/30/2015	AMAZON COM INC	AMZN	25.80%	1
10/31/2017	AMAZON COM INC	AMZN	24.96%	1
11/30/2021	ARISTA NETWORKS INC	ANET	33.65%	0
4/30/2014	ANADARKO PETROLEUM CORP	APC	2.52%	1
7/31/2013	ACTIVISION BLIZZARD INC	ATVI	-44.22%	1
2/28/2017	ACTIVISION BLIZZARD INC	ATVI	59.39%	1
1/31/2022	ACTIVISION BLIZZARD INC	ATVI	16.37%	1
11/30/2020	AMERICAN EXPRESS CO	AXP	-1.73%	1
8/31/2011	BANK OF AMERICA CORP	BAC	-19.57%	1
4/29/2011	BIOGEN IDEC INC	BIIB	N/A	1
7/31/2014	BIOGEN IDEC INC	BIIB	N/A	1
7/31/2018	BIOGEN INC	BIIB	3.63%	1
10/31/2019	BIOGEN INC	BIIB	N/A	1
2/28/2020	BIOGEN INC	BIIB	22.69%	1
11/30/2020	BIOGEN INC	BIIB	35.08%	1
6/30/2021	BIOGEN INC	BIIB	41.37%	1
9/30/2022	BIOGEN INC	BIIB	N/A	1
5/29/2015	BROADCOM CORP	BRCM	38.09%	1
7/31/2015	CHUBB CORP	CB	-3.69%	0
1/31/2019	CELGENE CORP	CELG	23.64%	1
6/30/2015	CIGNA CORP	CI	2.44%	1
8/31/2010	SALESFORCE COM INC	CRM	-1.83%	1
11/30/2010	SALESFORCE COM INC	CRM	8.80%	1
8/31/2020	SALESFORCE COM INC	CRM	22.64%	1

Appendix S.I - continued

DATE	Company Name	Ticker Symbol	ASV	$\mathbb{1}_{mssg}$
1/31/2017	C S X CORP	CSX	24.12%	1
5/28/2021	DOORDASH INC	DASH	0.19%	1
1/31/2013	DELL INC	DELL	-0.79%	0
5/31/2022	DOLLAR TREE INC	DLTR	N/A	1
9/30/2020	DOCUSIGN INC	DOCU	28.68%	1
6/30/2021	DOCUSIGN INC	DOCU	8.66%	1
7/30/2021	DEXCOM INC	DXCM	N/A	0
10/31/2022	DEXCOM INC	DXCM	N/A	0
4/30/2012	EBAY INC	EBAY	-3.61%	1
2/28/2018	EBAY INC	EBAY	-6.77%	1
7/29/2022	ENPHASE ENERGY INC	ENPH	26.65%	1
10/31/2011	EL PASO CORP	EP	-11.84%	0
10/31/2012	FACEBOOK INC	FB	19.92%	1
7/31/2013	FACEBOOK INC	FB	-0.60%	1
1/31/2014	FACEBOOK INC	FB	-0.80%	1
1/29/2016	FACEBOOK INC	FB	-31.11%	1
4/29/2022	META PLATFORMS INC	FB	15.05%	1
2/28/2014	FOREST LABS INC	FRX	10.13%	0
7/30/2010	GENZYME CORP	GENZ	16.34%	1
4/30/2012	GILEAD SCIENCES INC	GILD	3.90%	1
11/30/2012	GILEAD SCIENCES INC	GILD	-9.22%	1
7/29/2011	GOOGLE INC	GOOG	11.57%	0
10/31/2013	GOOGLE INC	GOOG	27.09%	0
7/31/2015	GOOGLE INC	GOOG	6.85%	1
7/31/2015	GOOGLE INC	GOOGL	N/A	1
7/30/2021	H C A HEALTHCARE INC	HCA	6.14%	0
2/28/2013	HEINZ H J CO	HNZ	N/A	1
2/28/2013	HEWLETT PACKARD CO	HPQ	13.27%	0
5/31/2013	HEWLETT PACKARD CO	HPQ	9.20%	0
5/29/2015	HUMANA INC	HUM	N/A	1
8/31/2017	ILLUMINA INC	ILMN	56.96%	1
7/31/2018	ILLUMINA INC	ILMN	50.92%	1
1/31/2022	ILLUMINA INC	ILMN	N/A	1
8/31/2010	INTUIT INC	INTU	N/A	1
1/29/2016	KINDER MORGAN INC	KMI	65.89%	1
3/31/2015	KRAFT FOODS GROUP INC	KRFT	51.46%	1
10/29/2021	LUCID GROUP INC	LCID	N/A	1
10/31/2019	LAM RESH CORP	LRCX	N/A	1
5/31/2019	MERCADOLIBRE INC	MELI	10.01%	1
7/29/2011	MEDCO HEALTH SOLUTIONS INC	MHS	-0.54%	1
11/30/2020	MODERNA INC	MRNA	53.15%	1
8/31/2021	MODERNA INC	MRNA	27.33%	1
11/30/2021	MODERNA INC	MRNA	0.70%	1
12/31/2021	MARVELL TECHNOLOGY INC	MRVL	26.29%	1
10/31/2011	MORGAN STANLEY DEAN WITTER & CO	MS	-3.15%	1
6/28/2019	MICRON TECHNOLOGY INC	MU	0.77%	1

Appendix S.I - continued

DATE	Company Name	Ticker Symbol	ASV	$\mathbb{1}_{mssg}$
1/31/2014	NETFLIX INC	NFLX	45.32%	1
1/30/2015	NETFLIX INC	NFLX	32.07%	1
4/30/2015	NETFLIX INC	NFLX	49.76%	1
7/31/2015	NETFLIX INC	NFLX	96.75%	1
10/31/2016	NETFLIX INC	NFLX	29.51%	1
7/31/2017	NETFLIX INC	NFLX	39.95%	1
9/30/2014	NIKE INC	NKE	17.55%	1
6/30/2021	NIKE INC	NKE	20.46%	1
11/30/2016	NVIDIA CORP	NVDA	40.46%	1
5/31/2017	NVIDIA CORP	NVDA	37.21%	1
8/31/2011	NEWS CORP	NWSA	N/A	1
12/31/2021	ORACLE CORP	ORCL	13.12%	1
8/31/2021	PALO ALTO NETWORKS INC	PANW	N/A	1
8/31/2010	PRICELINE COM INC	PCLN	N/A	1
8/31/2015	PRECISION CASTPARTS CORP	PCP	2.32%	1
10/30/2020	PINTEREST INC	PINS	4.83%	1
11/30/2020	PALANTIR TECHNOLOGIES INC	PLTR	N/A	1
1/29/2021	PALANTIR TECHNOLOGIES INC	PLTR	N/A	1
4/30/2019	QUALCOMM INC	QCOM	11.11%	1
10/31/2016	REYNOLDS AMERICAN INC	RAI	2.52%	1
5/28/2021	ROBLOX CORP	RBLX	N/A	1
11/30/2021	ROBLOX CORP	RBLX	N/A	1
9/30/2022	REGENERON PHARMACEUTICALS INC	REGN	-0.30%	0
10/31/2018	RED HAT INC	RHT	28.88%	1
1/31/2022	RIVIAN AUTOMOTIVE INC	RIVN	N/A	1
7/30/2021	ROKU INC	ROKU	-3.77%	1
4/30/2014	SPRINT CORP NEW	S	N/A	0
7/29/2016	SPRINT CORP NEW	S	N/A	0
5/31/2019	SPRINT CORP NEW	S	2.00%	0
2/28/2020	SPRINT CORP NEW	S	1.32%	0
6/28/2013	SPECTRA ENERGY CORP	SE	N/A	0
9/30/2016	SPECTRA ENERGY CORP	SE	-1.25%	0
4/30/2020	SNAP INC	SNAP	16.74%	1
10/30/2020	SNAP INC	SNAP	1.08%	1
7/30/2021	SNAP INC	SNAP	-0.69%	1
2/28/2022	SNAP INC	SNAP	6.38%	1
12/31/2021	SNOWFLAKE INC	SNOW	47.91%	1
8/31/2022	SNOWFLAKE INC	SNOW	2.31%	1

Appendix S.I - continued

DATE	Company Name	Ticker Symbol	ASV	$\mathbb{1}_{mssg}$
2/28/2022	BLOCK INC	SQ	12.77%	1
6/29/2018	SEMPRA ENERGY	SRE	4.40%	0
8/30/2019	TARGET CORP	TGT	24.99%	1
8/30/2013	TESLA MOTORS INC	TSLA	78.55%	1
2/28/2014	TESLA MOTORS INC	TSLA	63.66%	1
8/31/2018	TESLA INC	TSLA	91.88%	1
10/31/2018	TESLA INC	TSLA	54.76%	1
10/31/2019	TESLA INC	TSLA	14.61%	1
2/28/2020	TESLA INC	TSLA	69.19%	1
3/31/2021	TESLA INC	TSLA	23.45%	1
10/29/2021	TESLA INC	TSLA	28.30%	1
1/31/2022	TESLA INC	TSLA	20.74%	1
11/30/2020	TRADE DESK INC	TTD	13.35%	1
11/30/2021	TRADE DESK INC	TTD	11.35%	1
8/31/2022	TRADE DESK INC	TTD	21.74%	1
7/31/2014	TWITTER INC	TWTR	N/A	1
2/27/2015	TWITTER INC	TWTR	38.42%	1
7/31/2014	TIME WARNER INC NEW	TWX	22.49%	0
3/31/2020	UBER TECHNOLOGIES INC	UBER	-1.62%	1
6/30/2011	VISA INC	V	-8.39%	1
1/31/2013	VALERO ENERGY CORP NEW	VLO	32.00%	0
5/31/2022	VMWARE INC	VMW	N/A	0
3/31/2017	VERTEX PHARMACEUTICALS INC	VRTX	45.55%	1
7/31/2017	VERTEX PHARMACEUTICALS INC	VRTX	31.32%	1
6/29/2018	VERTEX PHARMACEUTICALS INC	VRTX	19.32%	0
9/30/2015	WESTERN DIGITAL CORP	WDC	2.53%	1
6/30/2014	WILLIAMS COS	WMB	21.93%	1
6/30/2015	WILLIAMS COS	WMB	2.52%	1
9/30/2020	ZOOM VIDEO COMMUNICATIONS INC	ZM	27.14%	1

Table S.II
Monthly analysis of maxret and future stock returns of large-cap portfolios

For each portfolio formation month, stocks are independently sorted into small- and large-cap groups, as well as into MAXRET deciles. Large-cap stocks are defined as those with market capitalizations above the NYSE median size breakpoint in that month. This table reports the average monthly returns of MAXRET-sorted portfolios comprising large-cap stocks. MAX events are defined in Table 1, and MAXRET refers to the return on the MAX day. If a stock experiences multiple MAX events in a given month, MAXRET is set to the highest of those returns. Columns 1−10 report value-weighted averages (in percentage points) of the following variables: excess return (RET−RF), the Fama-French-Carhart-Pastor-Stambaugh alpha (FFCPS), and the Fama-French five-factor alpha (FF5). The column labeled "High−Low" shows the difference between Decile 10 (High) and Decile 1 (Low) for each variable. Excess market returns and the size, book-to-market, momentum, profitability, and investment factors are from Kenneth French's data library; the liquidity factor is from Lubos Pastor's data library. Newey-West adjusted t-statistics are reported in parentheses. The sample period spans July 1963 through December 2022. *$p <$ .1; **$p <$ .05; *$p <$ .01.

	1 (Low)	2	3	4	5	6	7	8	9	10 (High)	High−Low
RET−RF	0.64	0.55	0.42	0.49	0.63	0.54	0.52	0.28	0.24	-0.23	-0.88**
	(4.12)	(3.34)	(2.42)	(2.52)	(2.70)	(2.22)	(1.82)	(0.84)	(0.65)	(-0.60)	(-2.41)
FFCPS	0.20	0.07	-0.10	-0.07	-0.03	0.00	-0.07	-0.33	-0.42	-0.82	-1.02***
	(1.82)	(1.12)	(-1.42)	(-0.89)	(-0.33)	(-0.02)	(-0.47)	(-2.07)	(-1.63)	(-2.44)	(-2.78)
FF5	0.13	-0.01	-0.15	-0.05	0.08	0.13	0.08	-0.08	-0.05	-0.51	-0.63*
	(1.31)	(-0.16)	(-2.68)	(-0.66)	(0.82)	(1.43)	(0.62)	(-0.52)	(-0.23)	(-1.67)	(-1.95)

Table S.III
State Social Media Privacy Laws

This table presents the implementation dates and statutory citations of state-level social media privacy laws that protect employees' use of social media from employer monitoring. These laws typically prohibit employers from requesting passwords, demanding access to personal accounts, or taking adverse actions based on employees' refusal to provide such access. Data are compiled from the Seyfarth Social Media Survey and additional research. The First Full Year column indicates the first complete calendar year in which the law was in effect.

State	Effective Date	Citation
Arkansas	2013-04-22	Password Protection Act (Act 1480, HB 1901)
California	2013-01-01	Labor Code 980 (AB 1844)
Colorado	2013-05-11	C.R.S. §8-2-127 (Social Media and the Workplace Act)
Connecticut	2015-10-01	Conn. Gen. Stat. §31-40x (PA15-6)
Delaware	2015-08-07	19 Del. C. §709A (HB109)
Hawaii	2021-06-07	Act 39 (HB 125) – UESOPPA
Illinois	2013-01-01	820 ILCS 55/10 (Right to Privacy in the Workplace Act)
Louisiana	2014-08-01	La. Rev. Stat. §51:1951-55 (Personal Online Account Privacy Protection Act)
Maine	2015-10-15	26 M.R.S. §616-619
Maryland	2012-10-01	Lab. & Emp. §3-712 (User Name & Password Privacy Protection Act)
Michigan	2012-12-27	MCL §37.271-278 (Internet Privacy Protection Act)
Montana	2015-04-23	Mont. Code §39-2-307
Nebraska	2016-07-21	Neb. Rev. Stat. §48-3501 et seq. (Workplace Privacy Act)
Nevada	2013-10-01	NRS §613.135
New Hampshire	2014-09-30	N.H. Rev. Stat. §275:74 et seq.
New Jersey	2013-12-01	N.J. Stat. §34:6B-5 to -10
New York	2024-03-12	N.Y. Lab. Law §201-i
Oklahoma	2014-11-01	40 Okla. St. §173.2
Oregon	2014-01-01	ORS §659A.330
Rhode Island	2014-06-30	R.I. Gen. Laws §28-56-1 et seq. (Social Media Privacy Act)
Tennessee	2015-01-01	Tenn. Code §50-1-1001 et seq. (Employee Online Privacy Act)
Utah	2013-05-14	Utah Code §34-48-201 et seq. (Internet Employment Privacy Act)
Vermont	2018-01-01	21 V.S.A. §495l
Virginia	2015-07-01	Va. Code §40.1-28.7:5
Washington	2013-07-28	RCW §49.44.200-.205
West Virginia	2016-06-10	W. Va. Code §21-5H-1 et seq.
Wisconsin	2014-04-10	Wis. Stat. §995.55

Table S.IV
Social interactions and lottery stock returns: Alternative lottery event definitions

This table examines the robustness of the results presented in Table 2 to alternative lottery event definitions. Panel A corresponds to a lottery event definition with a 10-day gap between a stock's daily return and the window of past returns. Specifically, a MAX event occurs when a stock's return on a given day (day 0) is equal to or greater than its return over the $[-32, -12]$ trading-day window. A stock experience a lottery event if it has a MAX event and the MAX-day return (MAXRET) falls within the top 10% of the cross-sectional distribution. Panel B corresponds to a path-independent lottery event definition, where a lottery event occurs if a stock's return on a given day ranks in the top 10% of the cross-sectional distribution of highest daily returns over the trailing 21 trading days. Since this alternatively-defined lottery events do not have an associated MAX event, it is not directly applicable to Table 2. We therefore estimate a modified version of this table by replacing MAXRET with an indicator for lottery events. See Table 2 for regression specification and additional variable definitions.

A. Lottery Event Definition, with a 10-day Gap

	Cumulative Returns		
	[1, 21]	[1, 42]	[1, 63]
	(1)	(2)	(3)
MAXRET $\cdot \mathbb{1}_{mssg}$	-0.103***	-0.109***	-0.177***
	(-3.26)	(-2.75)	(-3.37)
MAXRET	-0.078***	-0.109***	-0.097**
	(-2.71)	(-2.90)	(-1.98)
$\mathbb{1}_{mssg}$	0.002	0.000	0.003
	(0.47)	(-0.01)	(0.37)
BETA	0.001	0.003	0.002
	(0.45)	(0.80)	(0.52)
SIZE	-0.028***	-0.055***	-0.083***
	(-10.28)	(-11.01)	(-10.52)
BM	-0.001	0.000	0.004
	(-0.52)	(0.19)	(1.23)
IA	-0.002	-0.006*	-0.006
	(-1.05)	(-1.76)	(-1.32)
OP	0.032***	0.046***	0.053***
	(4.54)	(3.98)	(3.49)
MOM	-0.005*	-0.016***	-0.025***
	(-1.90)	(-4.34)	(-4.31)
REV	-0.020**	-0.038***	-0.063***
	(-2.30)	(-2.64)	(-3.48)
ILLIQ	-0.007	-0.013	-0.001
	(-0.43)	(-0.59)	(-0.02)
COSKEW	-0.014*	-0.023*	-0.024*
	(-1.71)	(-1.95)	(-1.96)
Obs.	436,243	432,103	426,104
Adj. R^2 (%)	11.64	16.93	18.16

B. Path-Independent Lottery Event Definition			
	Cumulative Returns		
	[1, 21]	[1, 42]	[1, 63]
	(1)	(2)	(3)
$\mathbb{1}_{lottery} \cdot \mathbb{1}_{mssg}$	-0.032***	-0.043***	-0.062***
	(-6.40)	(-6.91)	(-5.85)
$\mathbb{1}_{lottery}$	-0.015***	-0.020***	-0.021***
	(-5.25)	(-5.35)	(-4.56)
$\mathbb{1}_{mssg}$	-0.004*	-0.008**	-0.011**
	(-1.82)	(-2.27)	(-2.06)
BETA	0.002	0.003	0.004
	(1.09)	(1.16)	(1.05)
SIZE	-0.026***	-0.052***	-0.079***
	(-12.81)	(-12.33)	(-12.14)
BM	0.001	0.001	0.003
	(0.48)	(0.76)	(0.97)
IA	-0.003**	-0.007***	-0.010***
	(-2.44)	(-2.73)	(-2.79)
OP	0.034***	0.051***	0.059***
	(7.98)	(5.93)	(4.78)
MOM	-0.009***	-0.018***	-0.028***
	(-3.71)	(-5.34)	(-6.19)
REV	-0.016**	-0.033***	-0.046***
	(-2.51)	(-3.31)	(-4.14)
ILLIQ	0.014	0.031	0.038
	(1.27)	(1.50)	(1.33)
COSKEW	-0.005	-0.011	-0.017*
	(-0.96)	(-1.24)	(-1.74)
Obs.	9,011,326	8,926,038	8,839,161
Adj. R^2 (%)	10.65	14.58	17.15

Table S.V
Predicting lottery events: Alternative lottery event definitions

This table examines the robustness of the results presented in Table 3 to alternative lottery event definitions. Panel A corresponds to a lottery event definition with a 10-day gap between a stock's daily return and the window of past returns. Specifically, a MAX event occurs when a stock's return on a given day (day 0) is equal to or greater than it return over the $[-32, -12]$ trading-day window. A stock experience a lottery event if it has a MAX event and the MAX-day return (MAXRET) falls within the top 10% of the cross-sectional distribution. Panel B corresponds to a path-independent lottery event definition, where a lottery event occurs if a stock's return on a given day ranks in the top 10% of the cross-sectional distribution of highest daily returns over the trailing 21 trading days. See Table 3 for regression specification and additional variable definitions.

A. Lottery Event Definition, with a 10-day Gap

	Sample: All days $\mathbb{1}^{lottery}$		Sample: lottery event days Lottery-event day return	
	(1)	(2)	(3)	(4)
$\mathbb{1}^{preEvent}_{mssg}$	0.413***	0.475***	1.138**	1.515***
	(13.14)	(15.30)	(2.26)	(2.99)
$R^{preEvent}$	0.049	0.158	0.457	-1.695
	(0.23)	(0.77)	(0.42)	(-1.49)
BETA		0.034**		0.102
		(2.44)		(0.28)
SIZE		-0.318***		-3.081***
		(-21.31)		(-13.68)
BM		-0.014		-0.763***
		(-1.35)		(-3.17)
IA		-0.024*		-0.622**
		(-1.84)		(-2.37)
OP		-0.176***		-0.895
		(-4.68)		(-1.18)
MOM		-0.155***		-1.804***
		(-9.39)		(-8.33)
REV		-0.775***		3.054***
		(-10.42)		(3.17)
ILLIQ		0.017***		0.045
		(9.65)		(1.58)
COSKEW		0.004		-2.231*
		(0.13)		(-1.97)
Obs.	9,089,756	9,089,756	39,676	39,676
Adj. R^2 (%)	0.82	0.95	5.78	6.87

B. Path-Independent Lottery Event Definition

	Sample: All days $\mathbb{1}^{lottery}$		Sample: lottery event days Lottery-event day return	
	(1)	(2)	(3)	(4)
$\mathbb{1}^{preEvent}_{mssg}$	0.804***	0.860***	0.569*	0.963***
	(16.77)	(18.86)	(1.72)	(2.92)
$R^{preEvent}$	-0.305	-0.409	2.243***	1.687**
	(-0.98)	(-1.36)	(3.13)	(2.38)
BETA		0.056**		-0.050
		(2.48)		(-0.20)
SIZE		-0.518***		-2.110***
		(-23.32)		(-13.37)
BM		-0.034**		-0.435***
		(-2.14)		(-2.74)
IA		-0.041**		-0.330**
		(-2.00)		(-1.99)
OP		-0.361***		-0.579
		(-5.87)		(-1.26)
MOM		-0.242***		-1.332***
		(-8.01)		(-9.49)
REV		-0.042		-0.661**
		(-0.51)		(-2.09)
ILLIQ		0.030***		0.027
		(9.34)		(1.41)
COSKEW		0.020		-1.524**
		(0.43)		(-2.02)
Obs.	9,087,250	9,087,250	61,764	61,764
Adj. R^2 (%)	1.46	1.62	5.35	6.10

Table S.VI
Social interactions and retail herding in lottery stocks

This table examines the robustness of the results presented in Table 4 to alternative lottery event definitions. Panel A corresponds to a lottery event definition with a 10-day gap between a stock's daily return and the window of past returns. Specifically, a MAX event occurs when a stock's return on a given day (day 0) is equal to or greater than it return over the $[-32, -12]$ trading-day window. A stock experience a lottery event if it has a MAX event and the MAX-day return (MAXRET) falls within the top 10% of the cross-sectional distribution. Panel B corresponds to a path-independent lottery event definition, where a lottery event occurs if a stock's return on a given day ranks in the top 10% of the cross-sectional distribution of highest daily returns over the trailing 21 trading days. See Table 4 for regression specification and additional variable definitions.

A. Lottery Event Definition, with a 10-day Gap

	HR[0]	HR[1,5]	HR[6,10]
	(1)	(2)	(3)
$\mathbb{1}_{mssg}^{pre\text{Event}}$ (or $\mathbb{1}_{mssg}^{0}$)	0.039**	0.012*	0.005
	(2.29)	(1.81)	(1.00)
Return	0.035**	0.032***	0.005
	(2.16)	(4.84)	(1.12)
HR	1.031***	0.089***	0.029***
	(13.00)	(4.84)	(3.71)
USER	0.033***	0.014***	0.007***
	(7.65)	(12.43)	(5.51)
ΔUSER	0.323**	-0.004	-0.008
	(2.23)	(-0.50)	(-1.59)
AbVol	-0.023*	-0.005**	0.000
	(-2.00)	(-2.62)	(0.41)
AbNews	-0.016	-0.004*	-0.002*
	(-0.97)	(-1.99)	(-2.01)
EA	-0.023	-0.017***	-0.014***
	(-1.43)	(-3.69)	(-6.05)
Obs.	8,895	8,862	8,861
Adj. R^2 (%)	22.35	18.60	6.78

B. Path-Independent Lottery Event Definition			
	HR[0]	HR[1,5]	HR[6,10]
	(1)	(2)	(3)
$\mathbb{1}^{preEvent}_{mssg}$ (or $\mathbb{1}^{0}_{mssg}$)	0.028	0.021***	0.010*
	(1.44)	(3.02)	(1.74)
Return	0.039**	0.029***	0.004
	(2.72)	(4.11)	(0.78)
HR	0.973***	0.085***	0.030***
	(14.18)	(5.65)	(4.46)
USER	0.036***	0.014***	0.007***
	(7.99)	(11.39)	(6.14)
ΔUSER	0.332**	-0.001	-0.008
	(2.66)	(-0.11)	(-1.46)
AbVol	-0.011	-0.005**	0.000
	(-1.03)	(-2.37)	(0.12)
AbNews	-0.002	-0.003*	-0.003**
	(-0.11)	(-1.85)	(-2.32)
EA	-0.017	-0.018***	-0.015***
	(-0.82)	(-3.43)	(-5.14)
Obs.	10,116	10,643	10,642
Adj. R^2 (%)	19.25	17.92	6.99

Table S.VII
Social interactions and retail trading of lottery stocks: Alternative lottery event definitions

This table examines the robustness of the results presented in Table 5 to alternative lottery event definitions. Panels A and B correspond to a lottery event definition with a 10-day gap between a stock's daily return and the window of past returns. Specifically, a MAX event occurs when a stock's return on a given day (day 0) is equal to or greater than it return over the $[-32, -12]$ trading-day window. A stock experience a lottery event if it has a MAX event and the MAX-day return (MAXRET) falls within the top 10% of the cross-sectional distribution. Panels C and D correspond to a path-independent lottery event definition, where a lottery event occurs if a stock's return on a given day ranks in the top 10% of the cross-sectional distribution of highest daily returns over the trailing 21 trading days. See Table 5 for regression specification and additional variable definitions.

A. Lottery Event Definition, with a 10-day gap			
	Robinhood Sample: 06/2018−08/2020		
	OIB[0]	OIB[1,5]	OIB[6,10]
	(1)	(2)	(3)
$\mathbb{1}_{mssg}^{pre\text{Event}}$ (or $\mathbb{1}_{mssg}^{0}$)	2.230***	1.412**	0.583
	(3.43)	(2.63)	(1.11)
OIB	0.043	0.046***	0.034*
	(1.42)	(4.10)	(1.97)
Return	-0.738*	2.278***	0.572
	(-1.72)	(5.62)	(1.43)
HR	0.639**	1.168***	1.672***
	(2.17)	(3.16)	(3.19)
USER	-0.683**	1.130***	0.838***
	(-2.46)	(9.62)	(5.25)
ΔUSER	3.620	0.594	0.638**
	(1.33)	(1.20)	(2.14)
AbVol	0.005	-1.143***	-1.186***
	(0.01)	(-6.07)	(-4.84)
AbNews	-4.256***	-0.404	-0.055
	(-3.63)	(-1.46)	(-0.31)
EA	1.228	0.456	1.415**
	(1.02)	(0.86)	(2.24)
Obs.	8,895	9,074	9,073
Adj. R^2 (%)	0.98	2.82	1.84

B. Lottery Event Definition, with a 10-day Gap			
	StockTwits Sample: 01/2010−12/2022		
	OIB[0]	OIB[1,5]	OIB[6,10]
	(1)	(2)	(3)
$\mathbb{1}_{mssg}^{pre\text{Event}}$ (or $\mathbb{1}_{mssg}^{0}$)	-0.445	1.420***	1.338***
	(-1.22)	(5.72)	(5.92)
OIB	0.025*	0.034***	0.033***
	(1.88)	(7.04)	(7.06)
Return	-1.109***	1.120***	0.038
	(-3.05)	(3.34)	(0.14)
AbVol	-0.069	-0.700***	-0.524***
	(-0.23)	(-7.20)	(-6.62)
AbNews	-0.354	-0.635***	-0.625***
	(-0.81)	(-6.39)	(-6.33)
EA	-0.402	1.036***	1.978***
	(-0.88)	(3.42)	(6.07)
Obs.	50,297	52,648	52,585
Adj. R^2 (%)	7.02	8.30	6.79

C. Path-Independent Lottery Event Definition			
	Robinhood Sample: 06/2018−08/2020		
	OIB[0]	OIB[1,5]	OIB[6,10]
	(1)	(2)	(3)
$\mathbb{1}_{mssg}^{pre\text{Event}}$ (or $\mathbb{1}_{mssg}^{0}$)	2.141***	1.743***	0.441
	(4.01)	(3.38)	(1.27)
OIB	0.037	0.035***	0.028*
	(1.32)	(3.12)	(1.76)
Return	-0.619	2.456***	0.786*
	(-1.38)	(5.03)	(2.01)
HR	0.772**	0.824***	1.607***
	(2.07)	(2.85)	(3.25)
USER	-0.621***	1.137***	0.939***
	(-3.03)	(7.82)	(7.20)
ΔUSER	4.064	0.298	0.267
	(1.47)	(0.70)	(0.93)
AbVol	-0.429	-1.062***	-1.095***
	(-0.62)	(-7.12)	(-6.49)
AbNews	-3.612***	-0.306	0.006
	(-3.39)	(-1.22)	(0.03)
EA	1.419	0.183	1.314*
	(1.63)	(0.37)	(1.91)
Obs.	10,681	10,874	10,872
Adj. R^2 (%)	1.26	2.93	1.95

D. Path-Independent Lottery Event Definition			
	StockTwits Sample: 01/2010–12/2022		
	OIB[0]	OIB[1,5]	OIB[6,10]
	(1)	(2)	(3)
$\mathbb{1}^{pre\text{Event}}_{mssg}$ (or $\mathbb{1}^{0}_{mssg}$)	-0.432	1.448***	1.220***
	(-1.26)	(7.66)	(5.83)
Return	-1.020***	0.898**	-0.172
	(-3.46)	(2.59)	(-0.66)
OIB	0.042***	0.030***	0.027***
	(3.25)	(8.23)	(6.83)
AbVol	-0.062	-0.514***	-0.422***
	(-0.25)	(-7.09)	(-6.39)
AbNews	-0.536	-0.612***	-0.539***
	(-1.46)	(-6.81)	(-6.46)
EA	-0.385	0.853***	1.913***
	(-1.08)	(2.79)	(5.81)
Obs.	65,649	68,810	68,721
Adj. R^2 (%)	6.61	7.65	6.22