NBER WORKING PAPER SERIES

ATTENTION, SOCIAL INTERACTION, AND INVESTOR ATTRACTION TO LOTTERY STOCKS

Turan G. Bali
David Hirshleifer
Lin Peng
Yi Tang

Working Paper 29543
http://www.nber.org/papers/w29543

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
December 2021

We thank Hendrik Bessembinder, Ozgur Demirtas, Xi Dong, Joseph Engelberg (discussant), Jon Garfinkel, Danling Jiang, Lei Jiang, Patrick Konermann, Jens Kvarner, Eric Lie, Yueran Ma, Yiming Qian, Tyler Shumway (discussant), Johannes Stroebel, Paul Tetlock, Kamil Yilmaz, Jianfeng Yu, Xiaoyan Zhang, Dexin Zhou, 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 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. We thank Facebook for providing the anonymized SCI data. Lin Peng is grateful for data support she received from the Chinese University of Hong Kong ShenZhen while she was a visiting scholar. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Attention, Social Interaction, and Investor Attraction to Lottery Stocks
Turan G. Bali, David Hirshleifer, Lin Peng, and Yi Tang
NBER Working Paper No. 29543
December 2021
JEL No. D84,D91,G12,G14,G4,G41

ABSTRACT

We find that among stocks dominated by retail investors, the lottery anomaly is amplified by high investor attention (proxied by high analyst coverage, salient earnings surprises, or recency of extreme positive returns) and intense social interactions (proxied by Facebook social connectedness or population density near firm headquarters). Such stocks' lottery features attract greater Google search volume and retail net buying, followed by more negative earnings surprises and lower announcement-period returns. The findings provide insight into the roles of attention and social interaction in securities markets, and support the hypothesis that these forces contribute to investor attraction to lottery stocks.

Turan G. Bali
McDonough School of Business
Georgetown University
Between 37th and O Streets
Washington, D.C. 20057
turan.bali@georgetown.edu

David Hirshleifer
Marshall School of Business
University of Southern California
3670 Trousdale Pkwy
Los Angeles, CA 90089
and NBER
hirshlei@marshall.usc.edu

Lin Peng
Wasserman Department of
Economics and Finance
Baruch College
City University of New York
lin.peng@baruch.cuny.edu

Yi Tang
Gabelli School of Business
Fordham University
1790 Broadway
New York, NY 10019
ytang@fordham.edu

Related Working Papers (Bali) are available at https://sites.google.com/a/georgetown.edu/turan-bali
Related Working Papers (Hirshleifer) are available at https://sites.uci.edu/dhirshle/working-papers/
Related Working Papers (Peng) are available at https://sites.google.com/site/linpengbaruch/home
Related Working Papers (Tang) are available at https://sites.google.com/site/ytangweb/working-papers

There is strong evidence that investors—especially retail investors—are attracted to lottery stocks (stocks with positively skewed returns), and that this results in high valuations of such assets and low subsequent returns (see, e.g., Kumar 2009; Bali, Cakici, and Whitelaw 2011).[1] More recently, the arrival of fintech brokerage platforms with zero trading commissions has stimulated rising stock market participation by retail investors. In particular, Robinhood, the leading zero-commission broker, has used a variety of methods to gamify stock trading (Barber et al. 2021). For example, according to a *New York Times* article, "New members were given a free share of stock, but only after they scratched off images that looked like a lottery ticket."[2] These developments raise the questions of what drives retail investors' demand for lottery-type assets and how such demand influences asset prices.

A leading explanation for investor attraction to lottery stocks is that investors have nontraditional (imperfectly rational) preferences for portfolio skewness (Brunnermeier and Parker 2005; Brunnermeier, Gollier, and Parker 2007; Barberis and Huang 2008). This approach has typically assumed that investors perfectly know the return distributions of individual stocks. A crucial issue that has not been addressed in these models is how investors who are subject to attention constraints become *aware* of whether an asset is highly skewed. Furthermore, investor attraction to lottery stocks can be stimulated by social interaction even if investors have *no inherent preference for skewness* (Han, Hirshleifer, and Walden 2021).

Both approaches—a preference-based approach that is extended to allow for limited investor attention to lottery characteristics and a social interactions approach in which extreme high returns attract positive attention—imply that realized return outcomes as well as stock or investor characteristics affect the attraction of investors to lottery stocks. This paper therefore investigates how investor attention and social interactions affect the trading and pricing of lottery stocks.

[1] Lottery-like securities underperform in the U.S. equity market (Kumar 2009; Bali, Cakici, and Whitelaw 2011; Bali et al. 2017; Kumar, Page, and Spalt 2011, 2016; Wang, Yan, and Yu 2017; An et al. 2020), in international equity markets (Annaerta, DeCeustera, and Verstegena 2013; Walkshausl 2014; Barberis, Mukherjee, and Wang 2016; Zhong and Gray 2016; Carpenter, Lu, and Whitelaw 2016), and in other asset classes (Green and Hwang 2012; Boyer and Vorkink 2014). The effect is stronger for stocks with high retail ownership (Han and Kumar 2013; Bali et al. 2017; Lin and Liu 2018), but Agarwal, Jiang, and Wen (2019) also find evidence for an effect on mutual fund holdings.

[2] N. Popper, "Robinhood Has Lured Your Traders, Sometimes with Devastating Results," New York Times, July 08, 2020.

Extensive theoretical and empirical literature has studied how investor attention affects investors' information processing. Limited investor attention to public information can induce market underreactions to value-relevant news, whereas heavy investor attention can induce investor errors and price overreaction.[3] There are at least two different pathways by which investor attention and social interaction can affect the attraction to lottery stocks. First, even if investors have an inherent preference for positive skewness, they will still only select stocks on this basis if they are aware of and paying attention to these characteristics. Furthermore, more-extensive social interactions can help increase investors' awareness of positively skewed assets through word-of-mouth communication. Therefore, if investors inherently prefer high skewness, the arrival of news that makes the high skewness more visible and salient to investors will induce greater investor demand for and overpricing of skewness. Similarly, features of a firm's information environment that promote attention to lottery characteristics and greater social connectedness of the firm's investor base will have similar effects.

Second, even if investors have no inherent demand for skewness or volatility, investors may be attracted to volatile and positively skewed stocks by social interactions in which high returns are disproportionately reported and in which extremely high returns are highly salient (Han, Hirshleifer, and Walden 2021).[4] As a result, stocks with high volatility and skewness are in equilibrium overpriced and earn lower abnormal returns in the future. Because social interactions drive these effects, the model predicts that the attraction to and overpricing of skewness increases with the intensity of social interactions. Furthermore, these effects are predicted to be greatest for retail investors, who are most strongly subject to limited attention and the representativeness heuristic, two features that enhance the effect of social interactions on investor behavior.

[3]Theoretical and empirical studies of limited attention and market underreaction include Huberman and Regev (2001), Hou and Moskowitz (2005), Peng (2005), DellaVigna and Pollet (2007, 2009), Cohen and Frazzini (2008), and Hirshleifer, Lim, and Teoh (2009). Studies finding that heavy investor attention can induce price overreaction include Hirshleifer et al. (2004), Barber and Odean (2008), Tetlock (2011), and Gilbert et al. (2012). Kaniel and Parham (2017) find that media attention triggers a substantial increase in capital flows to mutual funds.

[4]In the model of Han, Hirshleifer, and Walden (2021), each "sender" (a current investor in the stock) can provide a report about the sender's recent return to another investor (the "receiver"). Owing to self-enhancing transmission bias, the higher a sender's return, the higher the probability the sender will report it to a receiver. Owing to the representativeness heuristic (itself a consequence of limited attention), receivers overextrapolate past returns. Also owing to the representativeness heuristic, receivers neglect the upward selection bias in the reports they receive (they hear more about high returns than low returns), which tends to make them overoptimistic about the stocks they hear about. This selection bias is more important when volatility is high, so receivers are on average attracted to high volatility stocks. Furthermore, receivers are more likely to pay attention to extreme return reports than intermediate ones. Therefore, among high volatility stocks, only highly extreme reports are likely to be reported and to be attractive to receivers. Hence, investors are also attracted to stocks with high skewness—even after controlling for volatility.

We therefore investigate whether an excessive attraction to lottery stocks is driven by investor attention and social interactions. We use the maximum daily return in the previous month (MAX) as the proxy for a stock's ex-ante lottery-like features (Bali, Cakici, and Whitelaw 2011). We hypothesize that retail investors are subject to stronger attentional biases and more exposed to the effects of social interactions. Our goal is to determine whether investor attention and social interactions are important drivers of individual investors' attraction to lottery stocks.

We apply three proxies to capture different aspects of investor attention. The first proxy, analyst following (CVRG), is associated with whether a firm has a high profile in public discussions. We expect that the extreme returns of firms that are in the public spotlight will tend to be disseminated widely. Consequently, more investors learn about a firm's lottery characteristics if that firm is in the public spotlight.

The second attention proxy is based on the magnitude of news events, measured by the absolute magnitude of the latest earnings surprises (|SUE|). As shown by Barber and Odean (2008) and Hirshleifer et al. (2008), firms that are in the news or have both positive and negative extreme earnings surprises tend to attract greater investor attention and buying. We therefore expect that firms with greater absolute earnings surprises are more likely to attract investor attention, thereby increasing investor awareness of the extent to which such firms have lottery characteristics.

The third attention proxy, RECENCY, captures the recency of a lottery event, and therefore reflects the dynamic decay of attention over time. The measure is motivated by the "recency effect" identified by studies in experimental psychology that show people tend to recall the most recent items best (Deese and Kaufman 1957; Murdock 1962). We therefore hypothesize that investor attention to extreme positive returns is higher the more recently these returns occurred.

The next two proxies are designed to capture the intensity of social interactions of a firm's investor base, namely, the population density (PD) and the Facebook Social Connectedness Index (SCIH) of the county where the firm's headquarters is located.

The use of PD measures is motivated by the finding that people in a more populated city have greater social connections and interactions (Hawley 2012; Bailey et al. 2018b) and the "home bias" phenomenon that shows investors tend to overweight local firms in their portfolios (see, e.g., Huberman 2001; Hong, Kubik, and Stein 2008). Thus, we expect that investors from more densely populated areas are also more likely to talk with their social network friends about their gains from investing in local stocks, especially those that exhibit lottery properties. News spreads through social interactions, and lottery stocks located in counties with greater social interactions are likely to attract greater investor attention and consequently greater demand for these stocks.

We also take advantage of a direct measure of social connectedness, the *Social Connectedness Index* (SCI), introduced by Bailey et al. (2018b). SCI measures investor social connectedness between U.S. counties based on friendship links on Facebook, the world's largest online social networking service. The enormous scale of Facebook's user base and the relative representativeness of its user body make the SCI a comprehensive measure of the geographic structure of the U.S. social networks.[5] We measure the social connectedness for each headquarters county (SCIH) as the aggregated SCI of the headquarters county with all other counties in the United States.

We first show that following an extreme daily return event, the increase in Google's abnormal search activities is markedly higher for stocks with greater analyst coverage, larger latest absolute earnings surprises, recent extreme positive return events, and stocks located in areas with high population density or Facebook social connectivity. This evidence suggests that our attention and social interaction proxies are indeed capturing investor attention to extreme high return events.

We next test the interplay between attention, social interactions, and the lottery anomaly. For stocks with high retail ownership, we find that the anomaly returns are higher for stocks with more analyst coverage, greater $|\text{SUE}|$, and more-recent MAX events, and is weaker for the low-attention stocks.

Specifically, the long-short MAX-sorted portfolios of stocks largely held by retail investors generate a monthly value-weighted alpha of -133 and -75 basis points for the high analyst coverage group and the low analyst coverage group, respectively. For stocks that have salient news, proxied by $|\text{SUE}|$, the alpha spread on the long-short MAX portfolio is -109 for the high $|\text{SUE}|$

[5] Facebook had 243 million active users in the United States and Canada as of the end of 2018. A 2018 survey indicated that 68% of U.S. adults report being Facebook users, with roughly three-quarters of those users visiting the site daily, and that users span a wide range of demographic groups (except for those 65 and older) (Smith and Anderson 2018). In addition, Duggan et al. (2015) and Bailey et al. (2018a, 2019, 2020) provide evidence that friendships observed on Facebook are a good proxy for real-world U.S. social connections.

group and -78 basis points for the low |SUE| group. In addition, the long-short MAX portfolio based on more-recent MAX events (high RECENCY group) produces a monthly alpha of -147 basis points, while the alpha for the stale MAX portfolio (low RECENCY group) is only -61 basis points.

Similarly, among stocks largely held by retail investors, the anomaly is more pronounced for stocks of firms headquartered in areas with more-active social interactions. For the high retail ownership stocks located in the top population density group, the long-short MAX portfolio generates a monthly value-weighted alpha of -135, while the corresponding alpha is -66 basis points for stocks located in the bottom PD group. Similarly, for the high retail ownership stocks located in the high SCIH region, the long-short MAX portfolio yields a value-weighted alpha of -151 basis points, while the corresponding alpha is only -56 basis points for stocks located in the low PD group. These results remain significant after controlling for a number of return predictors and socioeconomic variables simultaneously in Fama-MacBeth (1973) regressions.

So far, our finding that the lottery anomaly is most pronounced for retail stocks that attract high investor attention and social interactions is consistent with two explanations: skewness preference and social biased-belief. While several papers provide empirical evidence that supports the skewness preference explanation for the lottery anomaly,[6] the explanation based on social interactions and biased beliefs is new and has not yet been tested. If investor attraction to lottery stocks is driven by unwarranted optimism about future returns, as earnings news is realized, investors update their beliefs, resulting in a stock price correction. Therefore, the biased-beliefs hypothesis predicts that lottery stocks for which retail investors have extrapolative expectations have negative earnings announcement returns. Furthermore, negative announcement returns are expected to be more pronounced for stocks that attract greater attention and are located in areas with more-active social interactions. On the other hand, under the innate preference hypothesis, investors who are fully aware of the return distribution should be willing to accept a lower expected return in exchange for positive skewness. Such investors should not be disappointed at the lottery stock's earnings announcements in the future.

[6] See, e.g., Kumar (2009), Bali, Cakici, and Whitelaw (2011), Barberis, Mukherjee, and Wang (2016), Wang, Yan, and Yu (2017), and An et al. (2020) for empirical evidence of the preference-based models.

Our next set of analyses seek to differentiate the two mechanisms by investigating the extent to which investor attraction to lottery stocks is driven by biased-beliefs that derive from social interaction. We examine the subsequent realizations of individual firms' news and investors' reactions following a lottery event. Lottery stocks on average experience negative earnings surprises and large negative announcement returns in the subsequent month. Furthermore, such underperformance is mostly concentrated during the three days surrounding the earnings announcement. In addition, we find that retail investors' net purchases of a lottery stock are higher for stocks that are more likely to attract investor attention and for firms located in areas with high social connectedness. Taken together, our results are consistent with the predictions of Han, Hirshleifer, and Walden (2021) that investors' demand for lottery stocks is associated with investors forming biased expectations deriving from attention and social interactions.

Overall, our findings indicate that the general information environment, the availability of public news, and the recency of large positive return events affect retail investors' attention to lottery characteristics and contribute to the overvaluation of lottery stocks. Furthermore, we show that a higher intensity of social interactions contributes to stronger investor attraction to lottery stocks and greater overvaluation of such stocks. Our findings provide support for the hypothesis that *investor attention*, both direct and induced by *social interaction*, is a source of the lottery anomaly that is distinct from the direct effects of innate preferences.[7]

This paper provides the first empirical test of the two leading theories of investor attraction to lottery stocks. Among equity market anomalies, we focus on the lottery anomaly because it has plausible competing explanations based upon either the skewness preference theory, which is asocial, or a theory based on belief bias deriving from social interactions.[8] Our evidence helps to

[7]This result contrasts sharply with a common argument in previous literature that analyst following, as a proxy for the quality of the information environment, should be associated with *lower* mispricing (e.g., Hou and Moskowitz 2005; Hong, Torous, and Valkanov 2007; DellaVigna and Pollet 2009; Cohen and Frazzini 2008; Hirshleifer, Lim, and Teoh 2009; Hirshleifer, Hsu, and Li 2013; Bali et al. 2014). The key difference is that past studies focus on an advantage of high attention—that it reduces investor neglect of relevant signals such as earnings surprises or accruals. An alternative possibility that we examine here is that higher attention increases irrational investor attraction to lottery stocks.

[8]The skewness preference theory dates back to Arditti (1967) and Kraus and Litzenberger (1976), with several recent applications on the preference for lottery-like securities (i.e., Brunnermeier, Gollier, and Parker 2007; Barberis and Huang 2008; Kumar 2009; and Bali, Cakici, and Whitelaw 2011). On biased-belief and stock return anomalies in general, see Basu (1977), DeBondt and Thaler (1985), LaPorta and Shleifer (1997), Barberis, Shleifer, and Vishny (1998), and Daniel, Hirshleifer, and Subrahmanyam (1998); and for skewness, see Han, Hirshleifer, and Walden (2021).

distinguish asocial theories in which investors as isolated individuals have an inherent preference for skewness from theories in which social interaction attracts investor attention disproportionately to positively skewed stocks (regardless of whether investors are directly focused on skewness).

Survey evidence from Giglio et al. (2021) suggests that it is hard to predict when investors trade, but conditional on trading, belief changes affect both the direction and the magnitude of trades. Our approach helps identify when retail investors are more likely to be aware of a stock's lottery features, and therefore whether such awareness makes investors more likely to trade. Furthermore, conditional on trading, we find that the lottery features that grab investors' attention are those that lead them to revise upwards their expectations of returns. Hence our tests provide suggestive evidence regarding the questions of how attention, social interaction and beliefs affect the timing of investors' trades and, conditional on trading, the sizes of their trades.

Our paper also directly contributes to the more general literature on retail investors as a possible source of market return anomalies. Earlier studies show that retail investor attention is associated with overpricing and speculative trading (e.g., Barber and Odean 2008; Da, Engelberg, and Gao 2011; Yuan 2015; and Andrei and Hasler 2015). However, recent papers find that retail investors can actually contribute to market efficiency (Kelley and Tetlock 2013, 2017; Boehmer et al. 2021). Several contemporaneous studies focus on Robinhood investors. Welch (2021) documents aggregate buying activities by Robinhood investors during the pandemic and Barber et al. (2021) find that attention-induced herding by Robinhood investors is accompanied by large price movements and subsequent reversals.[9] Our findings that the speculative trading behavior of retail investors is associated with attention, is amplified by social interactions, and contributes to the lottery anomaly provide insights on the drivers of retail investor behavior and the sources of market inefficiencies.

Finally, the paper adds to the growing literature on the effect of social networks in financial markets. Models of rational social learning suggest that social interactions help disseminate valuable information (see, e.g., Ellison and Fudenberg 1995; Colla and Mele 2010; Ozsoylev and

[9] In addition, Ozik, Sadka, and Shen (2020), Glossner et al. (2020), and Eaton et al. (2021) study the effects of Robinhood investors on market liquidity.

Walden 2011).[10] On the other hand, social interactions can generate information cascades in which individuals do not make use of their private signals (Banerjee 1992; Bikhchandani, Hirshleifer, and Welch 1992), giving rise to free-riding incentives (Han and Yang 2013), or result in a Prisoner's Dilemma for the well-informed investors (Goldstein, Xiong, and Yang 2021). In addition, social interactions may amplify behavioral biases and spread inaccurate rumors, reducing information efficiency (see, e.g., Huberman 2001; Bailey, Kumar, and Ng 2011; Pool, Stoffman, and Yonker 2012; DeMarzo, Vayanos, and Zwiebel 2003; Kogan et al. 2019; Hirshleifer 2020). The lottery anomaly that we study provides an unique setting to test the bias-amplification hypothesis and our evidence is consistent with the efficiency-reduction effect of social interactions. These findings therefore suggest that the role of social network on financial markets are complex and nuanced.[11]

1. Data and Variable Definitions

We next describe the data sources and define the variables used in the empirical analyses. Our sample includes all common stocks traded on the NYSE, AMEX, and NASDAQ exchanges, covering the period from June 1963 through December 2017. We eliminate stocks with a price per share less than $5 or more than $1,000. The Center for Research in Security Prices (CRSP) provides the daily and monthly return and volume data. We adjust stock returns for delisting to avoid survivorship bias (Shumway 1997).[12]

Accounting variables are obtained from the merged CRSP-Compustat database. Analyst coverage data come from the Institutional Brokers' Estimate System (I/B/E/S) database and cover 1976–2017. The institutional ownership data are from Thompson 13F filings for 1980–2017. The excess market returns (MKT) and the size, book-to-market, momentum, profitability, and investment factors, namely, small minus big (SMB), high minus low (HML), momentum winner minus loser (UMD), robust minus weak (RMW), and conservative minus aggressive (CMA) are from

[10] The social interaction approach is supported by empirical studies on social networks and stock investments. See, e.g., Coval and Moskowitz 2001; Cohen, Frazzini, and Malloy 2008, 2010, Bernile, Kumar, and Sulaeman 2015; Pool, Stoffman, and Yonker 2015; Hong and Xu 2019; Da et al. 2019.

[11] Other recent papers using the newly available comprehensive Facebook data shows that social networks shape economic decisions and contribute to firms' access to institutional capital (Bailey et al., 2018a,b, 2020; Kuchler et al., 2021).

[12] Specifically, when a stock is delisted, we use the delisting return from CRSP, if available. Otherwise, we assume the delisting return is -100%, unless the reason for delisting is coded as 500 (reason unavailable), 520 (went to OTC), $551-573$, 580 (various reasons), 574 (bankruptcy), or 584 (does not meet exchange financial guidelines). For these observations, we assume that the delisting return is -30%.

Kenneth French's data library. The liquidity factor (PS) is from Lubos Pastor's data library. Unless otherwise stated, all variables are measured as of the end of the portfolio formation month (or month t) so that there is no look-ahead bias in our empirical analyses. We require a minimum of 24 monthly observations for variables computed from monthly data and a minimum of 15 daily observations for variables computed from daily data.

1.1. Key variables

Following Bali, Cakici, and Whitelaw (2011), we capture a stock's lottery feature with the stock's maximum daily return in the prior month (MAX).[13]

To characterize a stock's investor clientele, we define retail ownership (RHLD) as one minus its fractional institutional ownership. More specifically, we aggregate a stock's quarterly institutional ownership scaled by its total shares outstanding at the 8-character CUSIP level. We then merge the quarterly retail ownership variable with the CRSP data by CUSIP.[14]

We adopt three proxies that capture the effects of attention to a lottery stock. The first proxy measures the firm's prominence in public discussions. For this we use analyst following (CVRG), defined as the number of distinct earnings forecasts for a stock in the portfolio formation month. CVRG has been widely adopted by the aforementioned studies as a proxy for investor attention.[15] We expect analyst coverage to be associated with high attention for two reasons. First, analyst reports, recommendations, and forecasts are disseminated via the internet and the business media, and draw attention to the covered stock. Second, a stock of high interest to retail investors, such as Tesla, is more likely to receive analyst coverage.[16] Hence, individual investors are more likely to be aware of a lottery stock if more analysts cover it. To ensure that analyst coverage does not merely capture a size effect given its high correlation with firm size, we also examine an analyst coverage measure that is orthogonalized to size.

[13] As a robustness check, we also use the average of a stock's five highest daily returns in a month to proxy for the stock's ex-ante lottery-like characteristic. The results are very similar to those obtained from MAX.

[14] Following Cremers and Nair (2005), the quarterly institutional ownership is set to zero if missing in the database.

[15] Analysts tend to update forecasts in response to news. As such, the CVRG measured in month t may not proxy for investor attention, but rather for news releases. To mitigate this concern, we measure the CVRG over the past year. Our main findings remain unchanged, and they are available upon request.

[16] For our purposes, the direction of causality is not crucial, so long as coverage is positively correlated with attention.

The second and third proxy measure the magnitude and the recency of news events. The second proxy captures the salience of news and is motivated by Barber and Odean (2008) and Hirshleifer et al. (2008), who find that firms in the news and firms with both positive and negative earnings surprises tend to attract investor attention as reflected in buying and (in the tests of Barber and Odean) transient high returns. We therefore take the absolute value of a stock's unexpected earnings (|SUE|) to measure the saliency of a firm's news releases.[17] We expect a firm with greater $|\text{SUE}_{i,q}|$ to attract greater investor attention.

The third proxy, RECENCY, captures the dynamic decay of attention following a MAX event. It is defined as the inverse of one plus the number of trading days between the MAX day and the last trading day in the portfolio formation month. This measure is motivated by previous studies on how attention decays over time. For example, analyzing the collective attention to news stories by one million users of an interactive website, digg.com, Wu and Huberman (2007) find that the dynamics are well described by an attention-promoting novelty factor that decays over time. Similarly, Wang, Song, and Barabási (2013) and Higham et al. (2017) describe the decay of attention to paper and patent citations using exponential and log-normal functions. Candia et al. (2019) propose a universal biexponential function to fit the temporal decay of the attention received by cultural products (citation of academic articles and patents and the online attention received by songs, movies, and biographies). Hence, we conjecture that investor attraction to lottery events is higher for the more recent events (i.e., larger RECENCY).

The next two proxies are related to investor attractions to lottery stocks based on the intensity of social interactions of the county where the firm's headquarters is located. Investors tend to tilt their portfolios toward local firms, a phenomenon known as "home bias" (see, e.g., Huberman 2001; Hong, Kubik, and Stein 2008). These findings suggest that investors are more likely to talk with other people about their gains from investing in local stocks. Such a tendency increases for investors with greater social interactions, leading to more investor attraction to local stocks that have produced extreme positive returns.

[17] We follow Ball and Brown (1968) and Bernard and Thomas (1989, 1990) and define SUE as quarterly earnings surprises measured by the difference between the latest quarterly earnings per share after excluding extraordinary items (EPS) and the EPS four quarters ago, scaled by the standard deviation of quarterly earnings surprises over the past eight quarters.

The first proxy for the intensity of social interactions is the county-level population density (PD), based on the finding that people in a more populated city have greater social connections and interactions (Hawley 2012; Bailey et al. 2018b). The county-level population density (PD) is from the 1980, 1990, 2000, and 2010 U.S. Census. The decennial PD is linked to a firm's headquarters on the basis of the Federal Information Processing Standards (FIPS), which uses a five-digit coding system with the first two digits designating the state and the last three digits designating the county.

The second proxy of social interaction intensity is based on the Facebook Social Connectedness Index, which is available for April 2016. The SCI index, introduced by Bailey et al. (2018b), is a county-pair level measure that uses aggregated and anonymized information from the universe of friendship links among all Facebook users. More specifically, the Facebook SCI is calculated based on the total number of friendship ties as of April 2016 and is available for all pairs of 3,136 U.S. counties.[18] We define SCIH_i as the total connectedness of county i with all counties in the United States:

$$SCIH_i = \Sigma_{j=1}^{j} SCI_{i,j}, \tag{1}$$

where $\mathrm{SCI}_{i,j} \in [0,1]$ is the total number of friendship links between county i and county j, normalized by the maximum value of 1,000,000, which is assigned to the Los Angeles to Los Angeles county pair. We assign the county level SCIH to a firm based on its headquarters location and link the firm-level SCIH to stock information.[19]

In sum, PD and SCIH capture cross-sectional variations in investor attraction to lottery stocks that are driven by social interactions; CVRG captures variations in general attention as reflected in the firm's overall information environment; |SUE| reflects variations in attention driven by news arrival; and RECENCY uniquely captures the time dynamics of attention in relation to specific events. RECENCY also helps distinguish the effects of attention from rational effects of fundamental shocks, since in a frictionless rational setting, market adjustments should be instantaneous. All our proxies are of course imperfect, but by measuring several different aspects of attention

[18] For a detailed introduction of the data, see Bailey et al. (2018b).

[19] We assume that social interactions are relatively stable. We conduct a robustness check for the sample period of 2006−2017, during which the 2016 Facebook SCI data are more relevant; the results are robust.

(news-based triggers and the attention environment) and social interactions, our tests will suggest whether a clear general message emerges.

1.2. Control variables

We use a number of well-known cross-sectional return predictors as control variables in Fama and MacBeth (1973) regressions. Specifically, following Fama and French (1992), we estimate stock i's market beta (β^{MKT}) using its monthly returns over the prior 60 months if available (or a minimum of 24 months), 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 $t-1$.[20]

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). We define a stock's monthly co-skewness (COSKEW) following Harvey and Siddique (2000).[21] 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.[22] To control for the effect of post-earnings announcement drift, we follow Ball and Brown (1968) and Bernard and Thomas (1989, 1990) and use the standardized unexpected earnings (SUE) as quarterly earnings surprises measured by the difference between the latest quarterly earnings per share after excluding extraordinary items (EPS) and the EPS four quarters ago, scaled by the standard deviation of quarterly earnings surprises over the past eight quarters. Following Ang et al. (2006), the monthly idiosyncratic volatility (IVOL) is computed as the standard deviation of the return residuals from a regression

[20] Depending on availability, we use the redemption, liquidation, or par value (in that order) to estimate the book value of the preferred stock.

[21] $COSKEW_{i,t} = \frac{E[\varepsilon_{i,t} R^2_{m,t}]}{\sqrt{E[\varepsilon^2_{i,t}]}E[R^2_{m,t}]}$, where $\varepsilon_{i,t} = R_{i,t} - (\alpha_i + \beta_i R_{m,t})$ is the residual from the regression of the excess stock return ($R_{i,t}$) against the contemporaneous excess return on the CRSP value-weighted index ($R_{m,t}$) using the monthly return observations over the prior 60 months. The risk-free rate is measured by the return on one-month Treasury bills.

[22] 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.

of excess daily returns of a stock on the CRSP value-weighted index and the daily size and book-to-market factors of Fama and French (1993).

1.3. Summary statistics

Panel A of Table 1 provides the time series averages of the cross-sectional descriptive statistics for the aforementioned variables. Focusing on the proxies for attention and social interaction, we find that an average stock in our sample has a MAX daily return of 5.73%, is covered by seven analysts, has a $|SUE|$ of 0.84, has its highest daily return in the middle of a month, has a headquarters population density of 4,890 people per square mile, and has a SCIH score of 0.37. The standard deviations of the attention and social interaction proxies reflect substantial cross-sectional variations.

Panel B of Table 1 presents the time series averages of the cross-sectional correlations of the variables used in the study. Analyst coverage (CVRG) is highly positively correlated with firm size (SIZE) with an average cross-sectional correlation of 0.48, and is highly negatively correlated with retail ownership (RHLD), with an average cross-sectional correlation of -0.43. Furthermore, the PD and the SCIH are positively correlated, with an average cross-sectional correlation of 0.24 for 1976–2015,[23] indicating that people in more populated cities are more socially connected to the rest of the country, and that SCIH does capture a social interaction component that goes beyond PD.

2. The Lottery Anomaly and the Investor Clientele Effect

We first replicate prior findings on the lottery anomaly of Bali, Cakici, and Whitelaw (2011). For each month, we sort stocks into decile portfolios by the lottery feature proxies and report the average one-month-ahead value-weighted portfolio returns for the period of July 1963 through December 2017.[24] As shown in Table A1, the differences in the alphas between the high-MAX

[23]The average cross-sectional correlation between PD and SCIH for the recent decade of 2006–2015 is much stronger, at 0.37.

[24]We report the value-weighted average monthly excess returns and the corresponding risk-adjusted returns (alphas) relative to the Fama-French-Carhart-Pastor-Stambaugh model (FFCPS) and the Fama-French five-factor model (FF5), respectively. Newey-West 1987 t-statistics are given in parentheses. The FFCPS alphas are computed as the intercept from the regression of the value-weighted excess portfolio returns on a constant, the excess market return (MKT), the size factor (SMB), the book-to-market factor (HML), the momentum factor (MOM), and the liquidity risk factor (LIQ), following Fama and French (1993), Carhart (1997), and Pastor and Stambaugh (2003). The FF5

and low-MAX portfolios are negative and highly significant for the raw returns as well as the abnormal returns, indicating that stocks in the highest MAX decile generate lower risk-adjusted returns by 50 to 91 basis points per month compared to stocks in the lowest MAX decile. Furthermore, the negative alpha spread between high-MAX and low-MAX stocks is driven by the underperformance by high-MAX stocks.

Having established that the lottery anomaly remains strong and robust, we next turn to the analysis of its clientele effects. Both the preference-based and the social-interaction-based theories imply that lottery effects will be more pronounced for retail investors because retail investors are more attention-constrained and hence are more likely to be attracted to lottery stocks with better information environments or are subject to salient news and ignore less visible stocks. Retail investors are also more susceptible to representative heuristics and more likely to be influenced by the "word-of-mouth" effects of a social network. Thus, we expect the lottery anomaly to be stronger for stocks that are heavily held by retail investors.

At the end of each month, all stocks in the sample are grouped into three portfolios using the tercile breakpoints (30%/40%/30%) based on their retail ownership (RHLD). Stocks are also independently sorted into quintile portfolios based on MAX. The intersections of the three RHLD groups and the five MAX groups generate 15 value-weighted portfolios. Table 2 presents the FF5 alpha for each of the 15 portfolios. The last row reports the FF5 alpha spread between the high-MAX and low-MAX quintiles within each RHLD group. We find that the lottery anomaly returns are very high among stocks with high retail ownership but insignificant among stocks with low retail ownership. Specifically, the FF5 alpha spread is -84 basis points per month with a t-statistic of -3.48 for stocks in the high-RHLD group, whereas the FF5 alpha spread is insignificant for stocks in the low-RHLD group, at 16 basis points per month ($t = 0.89$). Overall, these results indicate that, among stocks with high retail ownership, the lottery anomaly is strong, but for stocks with low retail ownership, the lottery anomaly does not exist.[25]

alphas are computed with respect to the five-factor model of Fama and French (2015) with MKT, SMB, HML, CMA (investment), and RMW (profitability) factors. Throughout the paper, we calculate t-statistics using the Newey-West 1987 procedure with six lags.

[25]We report the value-weighted averages of MAX for each of the 15 portfolios in Table A2 of the online appendix. We will use these values in later sections to compute the economic significance of the slope coefficients obtained from Fama-MacBeth regressions.

The evidence of an investor clientele effect for the lottery anomaly suggests that attentional and social mechanisms could play important roles in driving the attraction to lottery stocks. Next, we directly examine the effect of investor attention and social interactions on the lottery anomaly, especially for stocks with high retail ownership.

3. Investor Attention, Social Interactions, and the Lottery Anomaly

In this section, we investigate how investor attention and social interactions affect the lottery anomaly. As explained in the introduction, if investors have an inherent preference for positive skewness, the arrival of news or the general information and social environment that make the skewness of a stock more visible and salient to investors will make the demand for the stock and its overpricing more sensitive to its skewness.[26] Even without an innate preference for skewness, the social-interaction-based model of Han, Hirshleifer, and Walden (2021) predicts that the attraction of investors to skewness increases with the intensity of social interactions.

Therefore, we next investigate how investor attention to a stock and the social connectedness of the stock's headquarters locations are associated with the lottery anomaly.

3.1. Visibility, news, and attention to lottery stocks

We first validate that our visibility and news proxies (CVRG, |SUE|, RECENCY) capture investor attention to stocks with lottery features. We do so by examining Google search activities in response to large positive stock returns following Da, Engelberg, and Gao (2011), who show that Google search activities regarding a stock capture attention from retail investors. More specifically, we measure a stock's abnormal retail attention as the abnormal search volume (ASV) of the stock, which is the percentage change between Google's daily Search Volume Index (SVI) for a stock and its past 12-month median.[27] Note that due to Google's short sample period of

[26] In the preference-based setting, the direction of the theoretical prediction for the interaction between attention and extreme news arrival could be ambiguous. On one hand, higher general attention means that investors tend to casually monitor the stock on a regular basis, and when there is a large positive return shock, investors will notice, devote more attention to the stock, and recognize its lottery feature. On the other hand, higher general attention means that investors are already closely evaluating the stock and thus know about its lottery characteristics. Therefore, when there is an extreme return event, investors are not surprised. Hence their lottery demand is not affected much. We later investigate Google search activities for stocks to distinguish these alternative predictions.

[27] The SVI is a relative search popularity score, defined on a scale of 0 to 100, based on the number of searches for a term relative to the total number of searches for a specific geographic area and for a given period. We focus on searches made on weekdays in the U.S. market. We manually screen all tickers to select those that do not have

January 2005 through December 2014 and limited cross-sectional coverage that focuses more on large firms, we only use ASV to validate our attention proxies and do not use ASV directly for the analysis of anomaly returns.

Figure 1 compares the ASV following MAX events for stocks in the top tercile and the bottom tercile CVRG or |SUE| groups. The upper (lower) panel shows that stocks in the top tercile CVRG (|SUE|) group attract more ASV than those in the bottom tercile CVRG (|SUE|) group on the MAX day and the subsequent five trading days. For instance, on the MAX day, ASV is 37.8% for stocks in the highest CVRG quintile, while only 3.5% for stocks in the lowest one. Similarly, ASV is 25.7% for stocks in the highest |SUE| quintile, while it is 20.8% for stocks in the lowest one. This pattern suggests that the general information environment and salient news events tend to increase investor attention to lottery features.

To confirm that attention to MAX events indeed decays over time (with RECENCY), we examine the average daily abnormal Google search volume (ASV) of stocks in the highest value-weighted MAX quintile portfolio on the MAX day (day 0) and the subsequent 21 trading days. Figure 2 shows that the average ASV is the highest, at more than 30%, on day 0, and it decays monotonically in the following month. Therefore, consistent with our conjecture, RECENCY captures the dynamic dimension of attention that reflects the information processing of attention-constrained investors over time. This decay pattern is less likely to be driven by observable time-invariant firm-characteristics or explained by instantaneous reactions to information shocks. The dynamics of Google search activities is consistent with evidence in experimental psychology (Deese and Kaufman 1957; Murdock 1962), and with the finding of decaying attention to news stories in digg.com (Wu and Huberman 2007), suggesting a general pattern of attention decay after salient events.

3.2. Attention and the lottery anomaly

Having established that our visibility and news proxies capture investor attention, we formally investigate the relation between these proxies and the lottery anomaly returns. Given our results

a generic meaning (e.g., "GPS" for GAP Inc., "M" for Macy's) to ensure that the search results we obtain are truly for the stock and not for other generic items or firm products. To avoid potential spillover effects in attention due to recent events, we exclude the most recent 20 days in computing the average SVI. We also exclude weekends because the markets are closed and search activities are very low.

in Section 2 that the lottery anomaly is mostly present for high-retail-ownership stocks, we focus on a sample of stocks within the highest tercile of retail holdings (RHLD). We construct two-way independently sorted portfolios, based on an attention proxy using tercile breakpoints, and MAX using quintile portfolios.[28]

Table 3 presents the FF5 alphas for the 15 value-weighted portfolios sorted by MAX and each of the three attention proxies. The lottery anomaly is presented in the last row, that is, it is the return difference between the high-MAX and the low-MAX quintiles within each attention-based group. The first three columns show that the FF5 alpha spread between the high-MAX and low-MAX quintiles gradually decreases as we move from the low-CVRG to high-CVRG group. The monthly FF5 alpha spread is negative and very large in absolute magnitude, −133 basis points (t = −2.24), for stocks in the high-CVRG group. The monthly FF5 alpha spread is −75 basis points (t = −2.78) for stocks in the medium-CVRG group and −75 basis points (t = −2.76) for stocks in the low-CVRG group. This result suggests that investor attention associated with a stock's general information environment increases the lottery anomaly.

Analyst coverage tends to be strongly correlated with firm size (Bhushan 1989). To control for the influence of size on analyst coverage, we follow Hong, Lim, and Stein (2000) and use the orthogonal component of analyst coverage ($\text{CVRG}^{size\perp}$), measured by the residuals of monthly cross-sectional regressions of the natural logarithm of the analyst coverage on the natural logarithm of market value of equity (in million dollars). We then repeat the analysis in Table 3 using $\text{CVRG}^{size\perp}$ and find qualitatively similar results. Table A4 shows that, for retail stocks within the high-$\text{CVRG}^{size\perp}$ group, the FF5 alpha spread between quintile 5 and quintile 1 of the MAX-sorted portfolios is −151 basis points per month and highly significant. On the other hand, for retail stocks within the low-$\text{CVRG}^{size\perp}$ group, the corresponding FF5 alpha spread is negative, but statistically insignificant. This result shows that the relation between analyst coverage and the lottery anomaly is not driven by firm size.

Our second attention proxy is based on a salient event, namely the size of the latest earnings surprise, |SUE|. The middle three columns of Table 3 demonstrate that the lottery anomaly in-

[28]For completeness, we present the detailed results of the 45 portfolios formed based on three-way independent sorts by RHLD, attention, and MAX in Table A3 of the online appendix.

creases monotonically when moving from the low-|SUE| to the high-|SUE| group. The FF5 alpha spread is negative and economically large, −109 basis points per month ($t = -3.61$), for stocks in the high-|SUE| group (strong attention-grabbing stocks). The FF5 alpha spreads are negative but smaller in absolute magnitude for stocks in the medium-|SUE| and low-|SUE| groups. This result suggests that a salient news announcement about a stock increases investor attention to its lottery feature and therefore intensifies the lottery anomaly.

Our third attention proxy, RECENCY, takes advantage of the fact that attention to lottery features decays over time. The larger the RECENCY, the more recently the lottery event occured relative to the portfolio formation; hence, it is likely to be associated with greater investor attention. The last three columns of Table 3 show that the lottery anomaly returns increase when moving from the low-RECENCY (more decay and less investor attention) to the high-RECENCY (less decay and more investor attention) group. The FF5 alpha spread is negative and very large in absolute magnitude, −147 basis points per month ($t = -4.01$), for stocks in the high-RECENCY group. The FF5 alpha spreads for stocks in the medium- and low-RECENCY groups are smaller in absolute magnitude, specifically, −87 basis points ($t = -3.14$) and −61 basis points per month ($t = -1.91$), respectively. The positive relation between the recency of a lottery event and the lottery anomaly returns is consistent with the decaying of attention to the event over time.

Overall, these results indicate that among stocks held by individual investors, the lottery anomaly is larger for stocks that receive greater investor attention due to a better information environment, more salient news, or recent MAX events.

3.3. Social interactions and the lottery anomaly

In this subsection, we investigate how social interactions affect investor attraction to lottery stocks. As discussed in the introduction, both the preference-based and the social interaction-driven attraction to lottery stocks are predicted to increase with the intensity of social interactions. This implies greater overvaluation of the lottery characteristic and lower future returns.

Our proxies for the intensity of social interactions of a stock's investor base are the population density (PD) of the county of the firm's headquarters and the Social Connectedness Index (SCIH) of the headquarters. We hypothesize that investor attraction to lottery stocks is positively associated with the degree of social interactions in the county in which the firm's headquarters is located. To test this hypothesis, for each month, we independently sort stocks into three groups based on a measure of social interactions and quintiles based on MAX. We calculate the value-weighted average daily abnormal Google search volume (ASV) on the MAX day as our proxy for retail attention. Figure 3 presents the ASV responses to MAX events for stocks that exhibit strong lottery characteristics, that is, stocks within the highest MAX quintile.

The upper panel of Figure 3 compares the average daily ASV between stocks in the high-PD and low-PD groups. It shows that the average daily ASV on the MAX day is always higher for stocks in the high-PD group than those in the low-PD group. A similar pattern is also observed for the subsequent five trading days. More specifically, for stocks in the high-PD group (with the highest degree of social interactions), the ASV is 18.8% and 13.0% on the MAX day and the next trading day, respectively. The ASV stays at around 7% in the subsequent four trading days. On the other hand, for stocks in the low-PD group (with the lowest degree of social interactions), the ASV is 9.1% on the MAX day. The ASV diminishes to 4.8% on the next trading day and becomes negligible after five trading days.

The lower panel of Figure 3 compares the average ASV between stocks in the high-SCIH and low-SCIH groups. Similarly, stocks in the high-SCIH group receive more abnormal retail investor attention than those in the low-SCIH group. These results confirm our conjecture that extremely positive daily returns for stocks whose headquarters are located in more socially connected areas attract greater investor attention.

Next, we directly examine the effect of social interactions on the lottery anomaly. We follow a procedure similar to that in Section 3.2 and focus on a sample of stocks within the highest tercile of retail ownership (RHLD). Specifically, we construct two-way independently sorted portfolios, based on a proxy for social interaction (PD or SCIH) and MAX.[29]

Table 4 presents the FF5 alphas for each of the 15 value-weighted portfolios from the intersections of the three social-interaction-sorted portfolios and the five MAX-sorted portfolios. The results show that the lottery anomaly is more pronounced for retail stocks with headquarters lo-

[29] For completeness, we present the detailed results of the 45 portfolios, triple-sorted by RHLD, a social interaction proxy, and MAX in Table A5 of the online appendix.

cated in more-populated areas (high-PD group), and is smaller for those with headquarters located in less populated areas (low-PD group). The FF5 alpha spread between quintile 5 and quintile 1 of the lottery-sorted portfolios is -135 basis points per month ($t = -4.73$) for stocks in the high-PD group, and -66 basis points per month ($t = -2.55$) for stocks in the low-PD group.

Turning to our second social interaction proxy, SCIH, Table 4 shows that the lottery anomaly returns are greater for stocks with headquarters located in areas with higher Facebook social connectedness. The FF5 alpha spread between quintile 5 and quintile 1 of the MAX-sorted portfolios is -141 basis points per month ($t = -4.09$) for stocks in the high-SCIH group and it reduces in absolute value to -91 basis points per month ($t = -3.33$) for stocks in the low-SCIH group. These results suggest that investors are more attracted to the salient MAX events for stocks located in more socially connected areas and, as a result, these stocks experience greater lottery premia.

In sum, the findings suggest that lottery stocks of firms headquartered in areas with intense social interactions attract more investor attention and are subject to more-pronounced overvaluation and lower future returns.

4. Cross-Sectional Regressions

We have so far investigated the factors that influence the lottery anomaly at the portfolio level. Although the portfolio-level analysis has the advantage of being non-parametric in the sense that we do not impose a functional form on the relation between attention, lottery-like features, and future returns, it does not allow us to account for all of the control variables jointly. To test whether our results hold after simultaneously controlling for well-known predictors of stock returns, we examine the impact of investor attention on the lottery anomaly using Fama and MacBeth (1973) regressions.

4.1. Investor attention and the lottery anomaly

We begin our regression analysis by investigating the effect of the information environment and news on lottery stock returns. The baseline model is the monthly cross-sectional regression with the following econometric specification:

$$R_{i,t+1} \quad = \quad \lambda_{0,t} + \lambda_{1,t} MAX_{i,t} + \lambda_{2,t} X_{i,t} + \varepsilon_{i,t+1}, \tag{2}$$

where $R_{i,t+1}$ is the realized excess return on stock i in month $t+1$, $MAX_{i,t}$ is a proxy for stock i's lottery-like payoffs in month t, and $X_{i,t}$ is a vector of control variables for stock i in month t, including the market beta (β^{MKT}), firm size (SIZE), book-to-market ratio (BM), momentum (MOM), illiquidity (ILLIQ), co-skewness (COSKEW), standardized unexpected earnings (SUE), and idiosyncratic volatility (IVOL).

Columns 1 and 2 in Table 5 report the time series averages of the slope coefficients over the 654 months from July 1963 through December 2017 for the full sample of NYSE/AMEX/NASDAQ stocks. The univariate regression results reported in Column 1 show a negative and statistically significant relation between MAX and the cross-section of future stock returns.[30] Column 2 shows that the average slope coefficient of MAX remains negative and statistically significant after accounting for all control variables simultaneously. These results confirm a significantly negative relation between lottery-like payoffs and future returns at the individual stock level.

Next, we investigate the effect of firm visibility and news on the lottery anomaly at the firm level using Fama and MacBeth (1973) regressions. We run monthly cross-sectional regressions for the following modified specification:

$$\begin{aligned} R_{i,t+1} \quad = \quad & \sum_{\tau=H,M,L} \alpha_t^{\tau} D_{i,t}^{\tau} + \sum_{\tau=H,M,L} \beta_t^{\tau} (MAX_{i,t} \times ATNT_{i,t} \times D_{i,t}^{\tau}) \\ & + \lambda_{1,t} MAX_{i,t} + \lambda_{2,t} ATNT_{i,t} + \lambda_{3,t} RHLD_{i,t} + \lambda_{4,t} X_{i,t} + \varepsilon_{i,t+1}. \end{aligned} \tag{3}$$

[30] The average slope, $\overline{\lambda_1}$, from the monthly regressions of realized returns on MAX alone is -0.0709, with a t-statistic of -5.88. The economic magnitude of the associated effect is similar to that documented in Table A1 for the value-weighted univariate portfolios. The spread in average MAX between deciles 10 and 1 is 13.62% (= 15.14% − 1.52%). Multiplying this spread by the average slope yields an estimate of the monthly lottery anomaly return of 0.97%.

In addition to the variables defined in Equation (2), we include the following: $\text{ATNT}_{i,t}$ denotes an attention proxy; and $\text{D}^{H}_{i,t}$, $\text{D}^{M}_{i,t}$, and $\text{D}^{L}_{i,t}$ are dummy variables, equal to one if a stock's RHLD is in the top, the middle, or the bottom tercile, respectively, and zero otherwise. Thus, the regression analysis here complements the portfolio analysis in Table 3 by controlling for other well-known predictors of future returns.

Columns 3 and 4 of Table 5 report the Fama-MacBeth regression results where the investor attention proxy is the natural logarithm of analyst coverage (LNCVRG). Consistent with the results from the portfolio analysis in Section 3.2, we find that investor attention enhances the lottery anomaly, and the effect is strongest for stocks in the top retail ownership group. More specifically, for stocks with the highest retail ownership (i.e., stocks with D^{H}=1), the average coefficient on the triple interaction term, β^{H}, is highly significant after controlling for the competing predictors of stock returns (Column 4). The average coefficient of -0.0387 implies that for a portfolio short-selling retail stocks in the low-MAX quintile and buying retail stocks in the high-MAX quintile, a one standard deviation increase of 0.99 in LNCVRG (see Table 1) increases the magnitude of the long-short portfolio returns by 44 basis points per month.[31] The average slope coefficients of the interaction terms are smaller and become insignificant for stocks in the medium and bottom retail holding groups, consistent with our findings in Section 2 that the lottery anomaly is only prevalent for retail-dominated stocks.

Columns 5−6 and 7−8 in Table 5 report the results for investor attention measured with |SUE| and RECENCY, respectively. The results show that the lottery anomaly is significantly stronger for retail stocks experiencing greater earnings surprises and more-recent MAX events. The average slope coefficient on the interaction between MAX, |SUE|, and D^{H} is -0.0418 after controlling for the full set of known return predictors (Column 6). Both are significant at the 1% level. In economic terms, a one standard deviation increase of 0.61 in |SUE| (see Table 1) increases the magnitude of the long-short lottery portfolio returns by 30 basis points per month ($= 0.0418 \times 0.61 \times 11.61\% = 0.30\%$). Similarly, the average coefficient on the interaction between MAX, RECENCY, and D^{H} is a significant -0.2077 after controlling for the full set of known return

[31] The increase of 44 basis points is calculated as the average slope coefficient on $MAX \times LNCVRG \times D^{H}$, 0.0387, multiplied by one standard deviation of LNCVRG, 0.99, multiplied by the net lottery exposure of the long-short portfolio, 11.61% (see Panel A of Table A2 in the online appendix).

predictors (Column 8). That is, a one standard deviation increase of 0.23 in RECENCY (see Table 1) increases the magnitude of the long-short lottery portfolio returns by 55 basis points per month ($= 0.2077 \times 0.23 \times 11.61\% = 0.55\%$).

Across all specifications, the coefficient estimates for control variables are largely in line with the findings of earlier studies. SIZE is negative and significant, consistent with Fama and French (1992). MOM is positive and significant, as in Jegadeesh and Titman (1993). SUE is positive and significant, while IVOL is negative and significant, consistent with Bernard and Thomas (1989, 1990) and Ang et al. (2006). Consistent with Harvey and Siddique (2000), COSKEW is negative and significant in Column 2, but loses its statistical significance in Columns 4, 6, 8. ILLIQ is significantly negative in Columns 2 and 4, but insignificant for the rest of the specifications.

Thus, the regression results are consistent with our portfolio results and suggest that the lottery anomaly is stronger for high-retail-ownership stocks that attract greater investor attention through a better information environment, more salient news releases, and recent lottery events.

4.2. Social interactions and the lottery anomaly

We next examine how social interactions affect the lottery anomaly. We run monthly cross-sectional regressions for the following specification:

$$\begin{aligned} R_{i,t+1} &= \sum_{\tau=H,M,L} \alpha_t^\tau D_{i,t}^\tau + \sum_{\tau=H,M,L} \beta_t^\tau (MAX_{i,t} \times SOCIAL_{i,t} \times D_{i,t}^\tau) \\ &\quad + \lambda_{1,t} MAX_{i,t} + \lambda_{2,t} SOCIAL_{i,t} + \lambda_{3,t} RHLD_{i,t} + \lambda_{4,t} X_{i,t} + \varepsilon_{i,t+1}. \end{aligned} \tag{4}$$

Note that Equation (4) is similar to Equation (3), where investor attention is now captured by one of the two social interaction proxies (SOCIAL): the population density (PD) and the social connectedness index (SCIH) of the county where a firm's headquarters is located. Table 6 presents the results.[32], [33]

[32]Panel B of Table 1 shows that retail ownership (RHLD) and firm size (SIZE) are highly negatively correlated with an average cross-sectional correlation of -0.14. For a robustness check, we orthogonalize retail ownership relative to firm size by regressing RHLD on the natural logarithm of market capitalization monthly. We define the high, medium, and low retail groups based on the regression residuals replacing the raw measure of retail ownership. We then replicate Equation (3). As shown in the first three columns of Table A6, the results based on the orthogonal component of RHLD remain unchanged.

[33]We further estimate Equations (3) and (4) using a sample of retail stocks (i.e., stocks in the top tercile retail ownership group or $D^H = 1$). As shown in Table A7 of the online appendix, the average slope coefficients on the inter-

In Columns 1 and 2 of Table 6, the social interaction proxy is PD, where Column 1 is the baseline regression and 2 controls for the full set of known return predictors. Consistent with the results from the portfolio-level analysis, we find that PD enhances the lottery anomaly and the effect is strongest for stocks in the top retail ownership group. For stocks with the highest retail ownership (i.e., stocks with D^H=1), the average coefficients on the interaction term, β^H, are -0.1395 in Column 1 and -0.1656 in Column 2 and both are statistically significant. In terms of economic significance, a one standard deviation increase of 13.39 in population density (see Table 1) translates into an increase in the lottery premium of 26 basis points per month (= $(-0.1656 \div 100) \times 13.39 \times 11.61\% = 0.26\%$), based on a long-short portfolio sorted by MAX.

The results based on SCIH are similar. For stocks with the highest retail ownership, the average coefficients on the interaction term are -0.0679 in Column 3 and -0.0554 in Column 4, and both coefficients are highly significant. In economic terms, after accounting for all control variables, a one standard deviation increase of 0.42 in SCIH (see Table 1) can be translated into an increase in the lottery anomaly return of 27 basis points ($= 0.0554 \times 0.42 \times 11.61\% = 0.27\%$) per month for a long-short portfolio sorted by MAX. In contrast, for stocks in the medium and bottom retail holding groups, the average coefficients of the interaction terms are smaller and become insignificant. In sum, our regression analyses corroborate the results of portfolio-level analysis and show that the lottery anomaly is greater for retail stocks from areas with more intense social interactions.

5. Preference, Beliefs, and Trading

So far, our findings that the lottery anomaly is most pronounced for retail stocks that attract high investor attention and are located in counties with intense social interactions are consistent with both the preference-based and the social biased-beliefs-based explanations. Earlier studies have explored the role of investor preferences,[34] whereas the new explanations of Han, Hirshleifer, and Walden (2021), which is based on social interaction and biased beliefs has not been tested. In

actions between MAX and proxies for investor attention (LNCVRG, |SUE|, and RECENCY) and social interactions (PD and SCIH), remain negative, and the magnitudes are highly comparable to those estimated using the full sample.

[34] See, for example, Brunnermeier and Parker (2005), Brunnermeier, Gollier, and Parker (2007), and Barberis and Huang (2008) for theoretical evidence, and Kumar (2009), Bali, Cakici, and Whitelaw (2011), Barberis, Mukherjee, and Wang (2016), Wang, Yan, and Yu (2017), and An et al. (2020) for empirical evidence.

this section, we investigate the extent to which our results are attributable to innate preferences or biased expectations.

We first examine whether our results can be explained by proxies of lottery preferences used in the prior literature. We then ask what types of investor beliefs drive their demand for lottery stocks by examining the subsequent realizations of firm news and market reactions.

5.1. Innate preference

Kumar (2009) finds that lottery stocks are particularly attractive to certain types of investors with an innate lottery preference: Catholics and people with low incomes and/or relatively lower education levels have a greater propensity to invest in lottery stocks. Our social interaction measures may be correlated with these demographic characteristics. For example, more-populated cities are likely to have lower household incomes because it may cost less to live in a city with public transportation.

To ensure that the impact of social interactions on the lottery anomaly is not due to these demographic characteristics, we collect county-level data on the percentage of the population who are Catholic (CATH), the percentage of the population who have a bachelor's degree or higher (EDU), and the county's median household income (MHI).[35] We then run monthly cross-sectional regressions for the following specification:

$$\begin{aligned} R_{i,t+1} = & \sum_{\tau=H,M,L} \alpha_t^\tau D_{i,t}^\tau + \sum_{\tau=H,M,L} \beta_t^\tau (MAX_{i,t} \times SOCIAL_{i,t} \times D_{i,t}^\tau) \\ & + \lambda_{1,t} MAX_{i,t} + \lambda_{2,t} SOCIAL_{i,t} + \lambda_{3,t} RHLD_{i,t} + \lambda_{4,t} X_{i,t} \\ & + \gamma_{1,t} CATH_{i,t} + \gamma_{2,t} EDU_{i,t} + \gamma_{3,t} MHI_{i,t} + \varepsilon_{i,t+1}. \end{aligned} \quad (5)$$

Table 7 reports the results. Columns 1 and 2 show that the average slope coefficients of the interaction terms are almost identical to those in Table 6: -0.1617 (t = -2.27) for $MAX \times PD \times$

[35] We obtain CATH from the U.S. Churches and Church Membership Survey in 1980, 1990, 2000, and 2010, which is available at the Association of Religion Data Archives' website, and EDU and MHI from the 1980, 1990, 2000, and 2010 U.S. Census. We link these decennial data to a firm's headquarters on the basis of the Federal Information Processing Standards (FIPS). Table 1 shows that at the county level, the Catholic population has a mean of 25.02% with a standard deviation of 14.59%, about one-third of the adult population has a bachelor's degree or higher with a standard deviation of 8.64%, and the median household income is \$40,720 with a standard deviation of \$9,540.

D^H and -0.0569 ($t = -3.53$) for $MAX \times SCIH \times D^H$. The coefficients for CATH, EDU, and MHI are all insignificant. These results indicate that these observable socioeconomic variables do not explain the impact of social interactions on the lottery anomaly, suggesting that our results are not likely to be a manifestation of investors' innate preferences for lottery-like securities.

5.2. Beliefs versus preferences

Next, we investigate the impact of investor beliefs on the lottery anomaly by examining the ex-post realizations of firm news and investors' reactions. As predicted by Han, Hirshleifer, and Walden (2021), investor attraction to lottery stocks is driven by unwarranted optimism about future returns. As earnings news is realized, investors update their beliefs, resulting in a stock price correction. Therefore, the biased belief hypothesis predicts that lottery stocks for which retail investors have extrapolative expectations have negative earnings announcement returns. Furthermore, negative announcement returns are expected to be more pronounced for stocks that attract greater attention and are located in areas with more-active social interactions. On the other hand, under the innate preference hypothesis, if investors are fully aware of the return distribution and are willing to accept a lower expected return in exchange for positive skewness, they should not be disappointed at the lottery stock's future earnings announcements.[36]

We proxy for the realizations of firm news with earnings announcements in the month subsequent to the MAX event month. We measure investors' reaction to the realization of earnings news with the cumulative market-adjusted return (CAR) for the three days surrounding the earnings announcement (see, e.g., Frazzini 2006; Kaniel et al. 2012). The biased-belief-based explanation thus predicts that the average CAR in month $t+1$ is more negative with increases in investor attention and social interactions, while the preference-based explanation would predict no abnormal CAR on average.

To test this prediction, we obtain the reported dates of quarterly earnings (the variable RDQ) from the quarterly CRSP/Compustat Merged database for the period of July 1974 through De-

[36]One may argue that investors with lottery preferences dynamically update their assessment of a stock's lottery feature, and therefore the disappointing news may lead to a downward adjustment of such an assessment. However, Bali, Cakici, and Whitelaw (2011) show that a stock's lottery characteristic is highly persistent: stocks in the top MAX decile have a 68% probability of remaining in the top three deciles in the next month. Therefore, we believe that the disappointing earnings news is more likely to trigger the correction of biased beliefs than the slow adjustment of assessed lottery features.

cember 2017. For each month t, we partition retail-dominated lottery stocks (i.e., stocks in the top MAX quintile group and the top tercile RHLD group) into three groups with tercile breakpoints based on an attention proxy (CVRG and RECENCY) or a social interaction proxy (PD and SCIH).[37] Panel A of Table 8 presents the average CAR across the lottery stocks in each value-weighted portfolio that makes earnings announcements in month $t+1$. The results show that the average CARs for lottery stocks in the top CVRG, RECENCY, PD, and SCIH portfolios are -1.54%, -0.64%, -0.91%, and -0.68%, respectively, and are statistically significant at the 1% level. The average CARs for the bottom counterpart portfolios are -0.61%, -0.01%, 0.05%, and 0.03%, respectively. Therefore, consistent with the prediction of the biased-expectation hypothesis, lottery stocks with high investor attention and social interactions underperform those with low investor attention and social interactions by a range of 63 to 96 basis points in the three days surrounding the earnings announcements.

As a placebo test, for each attention and social interaction portfolio constructed previously, we calculate the average market-adjusted returns for month $t+1$ after excluding the three days surrounding the earnings announcement (RET$-$CAR). Panel B of Table 8 shows that the average RET$-$CAR and the FF5 alpha for lottery stocks in the top attention or social interaction portfolio are in the range of -0.48% to 0.52% and are statistically insignificant.

Furthermore, for each previously constructed portfolio, we calculate the average standardized unexpected earnings (SUE) for earnings announced in month $t+1$ under the assumption that investors' biased beliefs are formed based on a seasonal random-walk model. Table 9 shows that the average SUE for lottery stocks in each top attention and social interaction portfolio, with no exceptions, is more negative than for those in the corresponding bottom portfolio. This result is also consistent with Engelberg, McLean, and Pontiff (2018), who find that the anomaly returns are significantly higher on earnings announcement days and other corporate news days and who argue that information arrival helps correct investors' biased expectations.

Overall, we find that the subsequent underperformance of retail-dominated lottery stocks is concentrated on the three days surrounding the realization of earnings news, and such underper-

[37] We do not consider |SUE| in this test because earnings news realized in the past quarter represents a distant memory and is a less relevant attention measure concerning one-quarter-ahead earnings news.

formance is more severe for lottery stocks that have more investor attention and social interactions. The evidence supports Han, Hirshleifer, and Walden (2021) in that investors have extrapolative expectations for stocks with large MAX returns, and such expectations are amplified by firm visibility, salient news and social interactions, but the investors are later disappointed with the subsequent earnings announcements.

5.3. Retail trading activities

Having established evidence that the lottery anomaly is associated with investors' overly optimistic beliefs about the stock's future returns, we examine retail trading activities following a MAX event. We conjecture that retail investors who are overly optimistic about lottery stocks' future returns are more likely to purchase than sell lottery stocks. Furthermore, as we have shown that investor attraction to lottery stocks increases with stock visibility, salience and recency of news, and social interactions, we expect retail order imbalances to increase with proxies for investor attention and social interactions.

Following Boehmer et al. (2021) (hereinafter BJZZ), we measure retail order imbalances by obtaining trades that occur off-exchange (i.e., with an exchange code equal to "D") for the period of January 2010 through December 2017 from the TAQ database.[38] For each day d, we define stock i's retail order imbalances based on share volume (OIBVOL) and the number of trades (OIBTRD) as follows:

$$OIBVOL_{i,d} = \frac{BVOL_{i,d} - SVOL_{i,d}}{BVOL_{i,d} + SVOL_{i,d}}, \tag{6}$$

$$OIBTRD_{i,d} = \frac{BTRD_{i,d} - STRD_{i,d}}{BTRD_{i,d} + STRD_{i,d}}, \tag{7}$$

where $BVOL_{i,d}$ and $SVOL_{i,d}$ are the number of stock i's shares bought and sold by retail investors on day d, respectively; and $BTRD_{i,d}$ and $STRD_{i,d}$ are the corresponding number of purchases and

[38] A transaction price in stock i at time s, $P_{i,s}$, is classified as a retail buy transaction if the fraction of a penny associated with $P_{i,s}$ is in the interval of (0.6, 1) and a retail sell transaction if in the interval of (0, 0.4). Following BJZZ, our order imbalance variables start in 2010, although data on subpenny improvement are available back to 2005. BJZZ find that in the initial few years before 2010, there is an upward bias in the subpenny trade data, which is possibly due to an increasing number of retail traders and brokerage firms' adopting subpenny improvement practices.

sales on the day, respectively. Table A8 of the online appendix shows that the summary statistics for our order-imbalance variables and BJZZ's are highly comparable.[39]

We define a MAX event if the maximum daily return of a stock belongs to the top MAX quintile of all stocks in a month. We then measure retail trading activities following MAX events relative to other information events by comparing them with the trading activities following earnings announcements, one of the most important firm news events. For each month, we compute the average differences in the retail order imbalances between the MAX event and the earnings announcements event. Panel A of Table 10 shows that the average OIBVOL and OIBTRD for the MAX event are significantly larger than those for the earnings event by 1.72% and 2.10%, respectively. These results imply that compared to earnings news, the lottery event is significantly more salient to retail investors and attracts excessively more retail buys than sells.

To test whether retail order imbalances for lottery stocks increase with investor attention and social interactions, we partition lottery stocks into three tercile groups based on an ascending sort of a proxy for investor attention (CVRG, |SUE|, and RECENCY) or social interactions (PD and SCIH). Panel B of Table 10 presents the time series averages of the cross-sectional mean of retail order imbalances for each portfolio. Consistent with our conjecture, the results show that the average retail order imbalances on the day following MAX are positive for stocks in the top attention or social interaction group and are statistically significant in most cases. On the other hand, the magnitudes of the order imbalances are smaller and statistically insignificant without any exception for stocks in the bottom attention or social interaction group.

Overall, our results suggest that not only are retail investors attentive to lottery stocks, but they act on their overly optimistic beliefs by engaging in active net buying of such stocks. Such trading behavior is more intensified for stocks with more visibility, salient news, and for those located in areas with greater social interactions. This evidence therefore support the view that the social and behavioral channels are important drivers of investor attraction to lottery stocks.

[39] The small differences are likely due to the different price screens applied in the two studies. We exclude stocks with a price per share less than $5 or more than $1,000. BJZZ exclude stocks with a per-share price less than $1. Consistent with the findings in BJZZ, retail sells are slightly more prevalent than buys. To mitigate the potential bias in retail order imbalances, we demean individual stocks' daily order imbalances by the cross-sectional mean of order imbalances.

6. Alternative Explanations

We have shown that the lottery anomaly is stronger for retail stocks that receive more investor attention and that are located in areas with intense social interactions. In this section, we investigate whether costly arbitrage, information supply, and market microstructure effects can provide complementary explanations to our main findings.

6.1. Costly arbitrage

The prior literature generally relies on stock characteristics such as firm age, size, analyst coverage, illiquidity, and idiosyncratic volatility to capture arbitrage costs (see, e.g., Amihud 2002; Pontiff 2006; Stambaugh, Yu, and Yuan 2015). On the other hand, stocks with high retail ownership tend to be young and have small market capitalization, low analyst coverage, high illiquidity, and high idiosyncratic volatility. Thus, retail ownership may not proxy for investor clientele but rather, arbitrage costs.

To test whether costly arbitrage explains our findings, we first construct an arbitrage cost index following a procedure employed by Stambaugh, Yu, and Yuan (2012, 2015). For each month, stocks in our sample are independently sorted into decile portfolios based on retail ownership (RHLD), market capitalization (SIZE), analyst coverage (CVRG), illiquidity (ILLIQ), idiosyncratic volatility (IVOL), and age, as measured by the number of months a stock is present on the CRSP database, in such an order that a higher portfolio rank is associated with more binding arbitrage costs. The arbitrage cost index (COST) is defined as the arithmetic average of the ranks of the six stock characteristics with a minimum of three of them available.

We re-estimate Equations (3) and (4) including COST as an additional control variable. Table 11 shows that for stocks with the highest retail ownership (i.e., stocks with $D^H = 1$), the average slope coefficients on the triple interaction term are significantly negative, and the magnitudes are highly comparable to those reported in Section 4. The average slope coefficients of the triple interaction term are smaller for stocks in the medium retail ownership group and are generally insignificant for those in the bottom retail ownership group. Hence, the results suggest that the impact of attention and social interactions on the lottery anomaly is not driven by arbitrage costs.

6.2. Information supply

Companies headquartered in areas with high population density are likely to be discussed by media and are more likely to be discussed following extremely high returns. To test whether our main findings are driven by the information supply channel, we obtain news coverage data spanning the period from January 2000 to December 2017 from Ravenpack News. We measure information supply by the number of relevant news reports (NEWS) from credible sources as defined by Ravenpack. We re-estimate Equations (3) and (4) including NEWS as an additional control variable.

Table 12 shows that for stocks with the highest retail ownership (i.e., stocks with $D^H = 1$), the average slope coefficients on the triple interaction term are significantly negative after controlling for NEWS, and the magnitudes are highly comparable to those reported in Section 4. Therefore, our main findings remain unchanged after accounting for the information supply channel.

6.3. Market microstructure effects

Earlier studies show that lottery stocks are relatively small, low-priced, less liquid, and have high idiosyncratic volatility, and hence are prone to microstructure effects.[40] In this subsection, we test whether accounting for the market microstructure effects dents our main findings.

We first examine whether the role of attention and social interactions in attracting retail investors to lottery stocks remains significant after removing microcaps. Following Hou, Xue, and Zhang (2020), for each month, we eliminate microcap stocks with market capitalizations smaller than the 20th NYSE size percentile and then replicate Tables 5 and 6.[41] Panel A of Table 13 shows that for stocks with the highest retail ownership (i.e., stocks with $D^H = 1$), the average slope coefficients of the triple interaction terms remain significantly negative, and the magnitudes are highly comparable to those estimated using the full sample. Therefore, our main finding holds after removing microcaps: that is, investor attention amplifies the lottery anomaly and the effect is strongest for stocks in the top retail ownership group.

[40] See, for example, Kumar (2009) and Bali, Cakici, and Whitelaw (2011).

[41] We obtain the monthly NYSE size percentiles from Kenneth French's online data library.

Then, following Asparouhova, Bessembinder, and Kalcheva (2010), we perform weighted least squares (WLS) estimations to address the issue of microstructure-related biases in regression estimations. More specifically, we estimate Equations (3) and (4) by scaling observed returns by the gross return on the same stock in the prior month. Panel B of Table 13 shows that for stocks in the top retail ownership group, the slope coefficients on the triple interaction term are significantly negative. These slope coefficients imply that for a portfolio short-selling retail stocks in the low-MAX quintile and buying retail stocks in the high-MAX quintile, a one standard deviation increase in investor attention or social interactions results in an increase in the average return on the long-short portfolio by a range of 29 to 52 basis points after accounting for the microstructure effects and all other control variables. The average slope coefficients on the triple interaction term are smaller for stocks in the medium retail ownership group and become insignificant for those in the bottom retail ownership group. These results are highly consistent with those from the OLS-based Fama-MacBeth regressions in Section 4.

7. Conclusions

Recent evidence suggests that investors are attracted to stocks with lottery characteristics, resulting in overpricing of such stocks and lower future returns. A leading explanation is that investors are perfectly aware of these lottery characteristics but nevertheless have inherent nontraditional preferences that induce demand for skewness (see e.g., Brunnermeier and Parker 2005; Barberis and Huang 2008).

Although plausible, such models rule out two important considerations. First, investors need to learn about the lottery characteristics of different stocks and allocate their limited attention to such updating. Consequently, proxies for attention should shift the demand for lottery stocks. Second, even without inherent preferences for skewness, extreme high returns can attract favorable investor attention to stocks, and the social transmission of the return news generates higher demand for lottery stocks (Han, Hirshleifer, and Walden 2021).

We investigate these issues by testing for the relationship between the overpricing of lottery stocks with proxies for investor attention and for the intensity of social interaction of a stock's local investor base. We first establish that, following an extreme positive return event, the increase in

Google's abnormal search activities is substantially higher when that return event is more recent. It is also substantially higher for stocks with greater analyst coverage, that have larger latest absolute earnings surprises, and whose firms are headquartered in areas with higher population density and Facebook social connectedness. These findings suggest that environmental factors and events that increase investor attention to a stock and the strength of social interactions by the stock's investors influence investor attraction to the stock's lottery features.

Turning to return predictability, we find that the negative relation between lottery characteristics and future returns is driven by the underperformance of lottery stocks with high retail participation, consistent with retail investors' being more likely to have limited attention, being more likely to be attracted by the salient features of lottery stocks, and being more susceptible to the effects of social interaction. More important, for such stocks, the lottery anomaly is more pronounced for stocks that receive more investor attention and whose firms are headquartered in areas with more-active social interaction.

We conduct further analysis to shed light on whether our findings are consistent with the innate preference mechanism or the social biased-beliefs explanations. We show that our results are not driven by socioeconomic variables that have been used in the literature to proxy for investor preferences and risk attitudes. More important, we find that the MAX events are followed by disappointing earnings announcements and large negative announcement returns. In addition, the underperformance of lottery stocks is mostly concentrated during the three days surrounding the earnings announcement and is insignificant for the other days. Furthermore, excessive retail buying of lottery stocks is especially high for stocks that are associated with high investor attention and for those from areas with high social connectedness. This evidence therefore suggests that the overvaluation of lottery stocks is associated with retail investors' extrapolative expectations that are intensified by attention and social interactions.

Our findings present the first empirical evidence supporting the hypothesis that investor attention and social interactions are important contributors to investor attraction to lottery stocks. This suggests that it would be fruitful to extend theories of irrational investor preferences for lottery stocks to allow for imperfect investor knowledge about lottery characteristics and for how environmental cues can draw attention to such characteristics. It further suggests that models based upon social interactions have promise as an alternative explanation for investor attraction to lottery stocks. These mechanisms are likely to be particularly important in the new era of increased online social interactions and the growth of zero-commission trading platforms.

References

Agarwal, V., L. Jiang, and Q. Wen. 2019. Why do mutual funds hold lottery stocks? *Journal of Financial and Quantitative Analysis* Forthcoming.

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.

Arditti, F. D. 1967. Risk and the required return on equity. *Journal of Finance* 22:19–36.

Asparouhova, E. N., H. Bessembinder, and I. Kalcheva. 2010. Liquidity biases in asset pricing tests. *Journal of Financial Economics* 96:215–37.

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., E. Dávila, T. Kuchler, and J. Stroebel. 2019. House price beliefs and mortgage leverage choice. *Review of Economic Studies* 86:2403–52.

Bailey, M., P. Farrell, T. Kuchler, and J. Stroebel. 2020. Social connectedness in urban areas. *Journal of Urban Economics* 118:103264–.

Bailey, W., A. Kumar, and D. Ng. 2011. Behavioral biases of mutual fund investors. *Journal of Financial Economics* 102:1–27.

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., 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., L. Peng, Y. Shen, and Y. Tang. 2014. Liquidity shocks and stock market reactions. *Review of Financial Studies* 27:1434–85.

Ball, R., and P. Brown. 1968. An empirical evaluation of accounting income numbers. *Journal of Accounting Research* 6:159–78.

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. 2021. Attention-induced trading and returns: Evidence from robinhood. Forthcoming.

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.

Barberis, N., A. Shleifer, and R. Vishny. 1998. A model of investor sentiment. *Journal of Financial Economics* 94:307–43.

Basu, S. 1977. Investment performance of common stocks in relation to their price-earnings ratios: A test of the efficient market hypothesis. *Journal of Finance* 32:663–82.

Bernard, V. L., and J. K. Thomas. 1989. Post-earnings-announcement drift: Delayed price response or risk premium. *Journal of Accounting Research* 27:1–36.

———. 1990. Evidence that stock prices do not fully reflect the implications of current earnings for future earnings. *Journal of Accounting and Economics* 13:305–40.

Bernile, G., A. Kumar, and J. Sulaeman. 2015. Home away from home: Geography of information and local investors. *Review of Financial Studies* 28:2009–49.

Bhushan, R. 1989. Firm characteristics and analyst following. *Journal of Accounting and Economics* 11:255–74.

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.

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

Brunnermeier, M. K., C. Gollier, and J. A. Parker. 2007. Optimal beliefs, asset prices, and the preference for skewed returns. *American Economic Review* 97:159–65.

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

Candia, C., C. Jara-Figueroa, C. Rodriguez-Sickert, A.-L. Barabási, and C. A. Hidalgo. 2019. The universal decay of collective memory and attention. *Nature Human Behaviour* 3:82–91.

Carhart, M. M. 1997. On persistence in mutual fund performance. *Journal of Finance* 52:57–82.

Carpenter, J., F. Lu, and R. Whitelaw. 2016. The real value of china's stock market. Working paper, New York University.

Cohen, L., and A. Frazzini. 2008. Economic links and predictable returns. *Journal of Finance* 63:1977–2011.

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

———. 2010. Sell-side school ties. *Journal of Finance* 65:1409–37.

Colla, P., and A. Mele. 2010. Information linkages and correlated trading. *Review of Financial Studies* 23:203–46.

Coval, J. D., and T. J. Moskowitz. 2001. The geography of investment: Informed trading and asset prices. *Journal of Political Economy* 109:811–41.

Cremers, M., and V. Nair. 2005. Governance mechanisms and equity prices. *Journal of Finance* 60:2859–94.

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

Da, Z., U. G. Gurun, B. Li, and M. Warachka. 2019. Investment in a smaller world: The implications of air travel for investors and firms. *Management Science* 67:417–35.

Daniel, K. D., D. Hirshleifer, and A. Subrahmanyam. 1998. A theory of over-confidence, self-attribution, and security market under- and overreactions. *Journal of Finance* 53:1839 – 1886.

DeBondt, W. F. M., and R. H. Thaler. 1985. Does the stock market overreact? *Journal of Finance* 40:793–808.

Deese, J., and R. A. Kaufman. 1957. Serial effects in recall of unorganized and sequentially organized verbal material. *Journal of Experimental Psychology* 54:180–7.

DellaVigna, S., and J. Pollet. 2007. Demographics and industry returns. *American Economic Review* 97:1667–702.

———. 2009. Investor inattention and Friday earnings announcements. *Journal of Finance* 64:709–49.

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

Duggan, M., N. B. Ellison, C. Lampe, A. Lenhart, and M. Madden. 2015. Demographics of key social networking platforms. Pew Research Center.

Eaton, G. W., C. T. Green, B. Roseman, and Y. Wu. 2021. Zero-commission individual investors, high frequency traders, and stock market quality. Working paper.

Ellison, G., and D. Fudenberg. 1995. Word-of-mouth communication and social learning. *Quarterly Journal of Economics* 110:93–125.

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.

———. 1993. Common risk factors in the returns of stocks and bonds. *Journal of Financial Economics* 33:3–56.

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

Fama, E. F., and J. MacBeth. 1973. Risk, return and equilibrium: Empirical tests. *Journal of Political Economy* 51:55–84.

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.

Giglio, S., M. Maggiori, J. Stroebel, and S. Utkus. 2021. Five facts about beliefs and portfolios. *American Economic Review* 111:1481–522. doi:10.1257/aer.20200243.

Gilbert, T., S. Kogan, L. A. Lochstoer, and A. Ozyildirim. 2012. Investor inattention and the market impact of summary statistics. *Management Science* 58:336–50.

Glossner, S., P. Matos, S. Ramelli, and A. F. Wagner. 2020. Where do institutional investors seek shelter when disaster strikes? evidence from covid-19. Working paper.

Goldstein, I., Y. Xiong, and L. Yang. 2021. Information sharing in financial markets. Working paper.

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

Han, B., D. Hirshleifer, and J. Walden. 2021. Social transmission bias and investor behavior. *Journal of Financial and Quantitative Analysis* Forthcoming.

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.

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

Hawley, Z. B. 2012. Does urban density promote social interaction? evidence from instrumental variable estimation. *Review of Regional Studies* 42:223–48.

Higham, K. W., M. Governale, A. B. Jaffe, and U. Zülicke. 2017. Fame and obsolescence: Disentangling growth and aging dynamics of patent citations. *Physical Review E* 95:042309–.

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

Hirshleifer, D., K. Hou, S. H. Teoh, and Y. Zhang. 2004. Do investors overvalue firms with bloated balance sheets? *Journal of Accounting and Economics* 38:297–331.

Hirshleifer, D., P.-H. Hsu, and D. Li. 2013. Innovative efficiency and stock returns. *Journal of Financial Economics* 107:632–54.

Hirshleifer, D., S. Lim, and S. H. Teoh. 2009. Driven to distraction: Extraneous events and underreaction to earnings news. *Journal of Finance* 64:2289–325.

Hirshleifer, D. A., J. N. Myers, L. A. Myers, and S. H. Teoh. 2008. Do individual investors cause post-earnings announcement drift? direct evidence from personal trades. *Accounting Review* 83:1521–50.

Hong, H., J. Kubik, and J. Stein. 2008. The only game in town: Stock-price consequences of local bias. *Journal of Financial Economics* 90:20–37.

Hong, H., T. Lim, and J. C. Stein. 2000. Bad news travels slowly: Size, analyst coverage, and the profitability of momentum strategies. *Journal of Finance* 55:265–95.

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. 2020. Replicating anomalies. *Review of Financial Studies* 33:2019–133.

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

Huberman, G., and T. Regev. 2001. Contagious speculation and a cure for cancer. *Journal of Finance* 56:387–96.

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

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.

Kaniel, R., and R. Parham. 2017. WSJ category kings — The impact of media attention on consumer and mutual fund investment decisions. *Journal of Financial Economics* 123:337–56.

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.

Kogan, S., S. Kogan, T. J. Moskowitz, T. J. Moskowitz, and M. Niessner. 2019. Fake news: Evidence from financial markets.

Kraus, A., and R. H. Litzenberger. 1976. Skewness preference and the valuation of risk assets. *Journal of Finance* 31:1085–100.

Kuchler, T., Y. Li, L. Peng, J. Stroebel, and D. Zhou. 2021. Social proximity to capital: Implications for investors and firms. *Review of Financial Studies* Forthcoming.

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.

LaPorta, R., and A. Shleifer. 1997. Good news for value stocks: Further evidence on market efficiency. *Journal of Finance* 52:859–74.

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

Murdock, Jr., B. B. 1962. The serial position effect of free recall. *Journal of Experimental Psychology* 64:482–8.

Newey, W. K., and K. D. West. 1987. A simple, positive-definite, heteroscedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55:703–8.

Ozik, G., R. Sadka, and S. Shen. 2020. Flattening the illiquidity curve: Retail trading during the covid-19 lockdown. *Journal of Financial and Quantitative Analysis* Forthcoming.

Ozsoylev, H. N., and J. Walden. 2011. Asset pricing in large information networks. *Journal of Economic Theory* 146:2252–80.

Pastor, L., and R. F. Stambaugh. 2003. Liquidity risk and expected stock returns. *Journal of Political Economy* 111:642–85.

Peng, L. 2005. Learning with information capacity constraints. *Journal of Financial and Quantitative Analysis* 40:307–29.

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. 2012. No place like home: Familiarity in mutual fund manager portfolio choice. *Review of Financial Studies* 25:2563–99.

———. 2015. The people in your neighborhood: Social interactions and mutual fund portfolios. *Journal of Finance* 70:2679–732.

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

Smith, A., and M. Anderson. 2018. Social media use in 2018. Pew Research Center.

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.

Tetlock, P. C. 2011. All the news that's fit to reprint: Do investors react to stale information? *Review of Financial Studies* 24:1481–512.

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

Wang, D., C. Song, and A.-L. Barabási. 2013. Quantifying long-term scientific impact. *Science* 342:127–32.

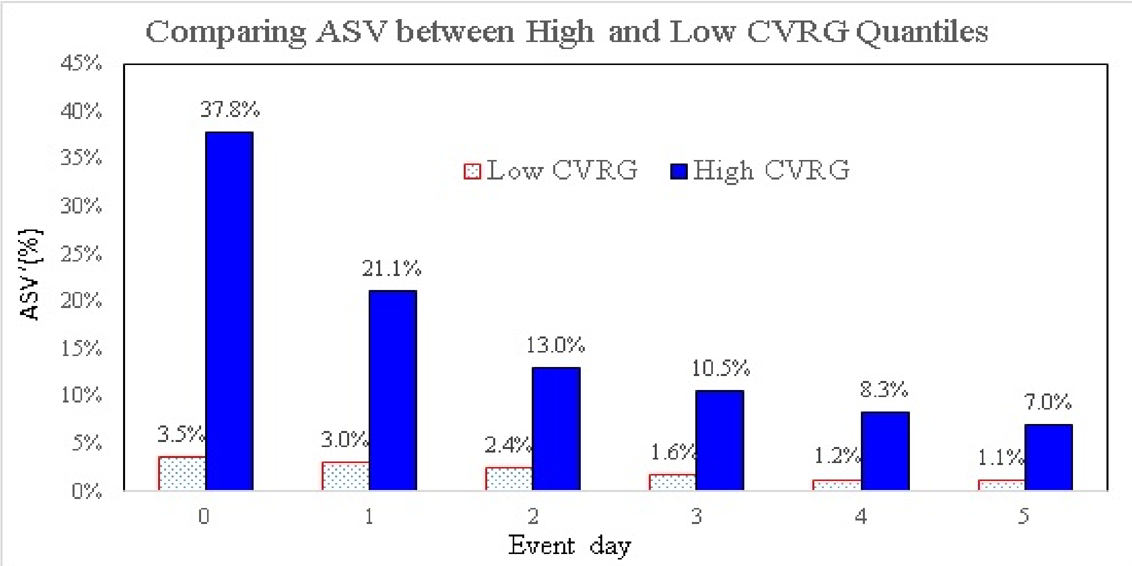
Wang, H., J. Yan, and J. Yu. 2017. Reference-dependent preferences and the risk-return tradeoff. *Journal of Financial Economics* 123:395–414.

Welch, I. 2021. The wisdom of the robinhood crowd. *Journal of Finance* Forthcoming.

Wu, F., and B. A. Huberman. 2007. Novelty and collective attention. *Proceedings of the National Academy of Sciences* 104:17599–601. ISSN 0027-8424. doi:10.1073/pnas.0704916104.

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.



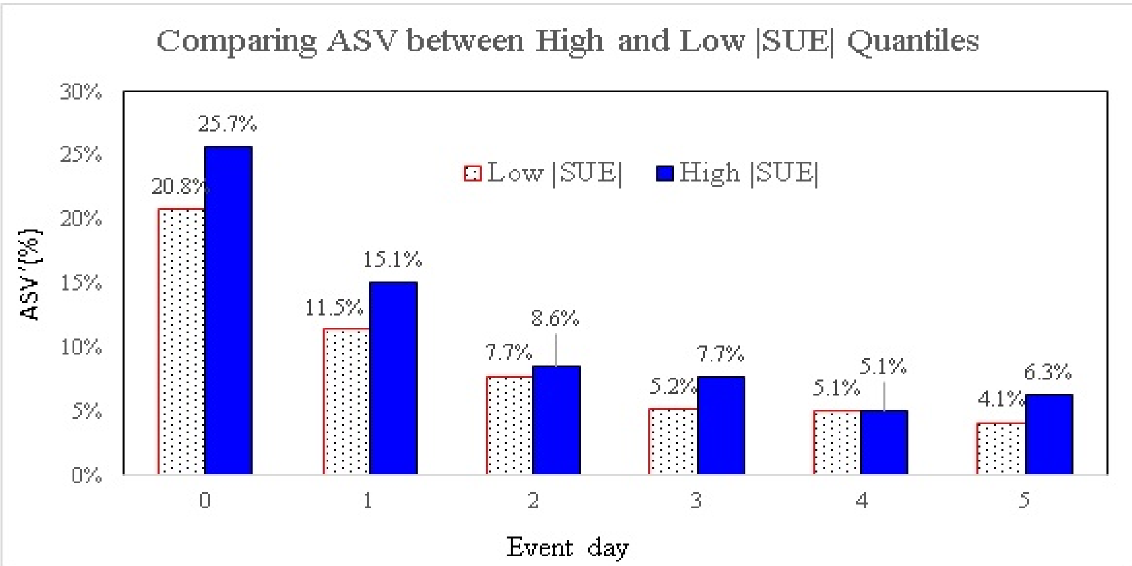


Figure 1. Comparing abnormal Google search volume between high and low attention groups. The solid (dotted) bars depict the average daily abnormal Google search volume (ASV) on the MAX day (day 0) and the subsequent five trading days of stocks within the highest value-weighted MAX quintile portfolio and the high (low) attention-based groups. The proxies for investor attention are analyst coverage (CVRG) in the upper panel and absolute value of the standardized earnings surprise (|SUE|) in the lower panel.

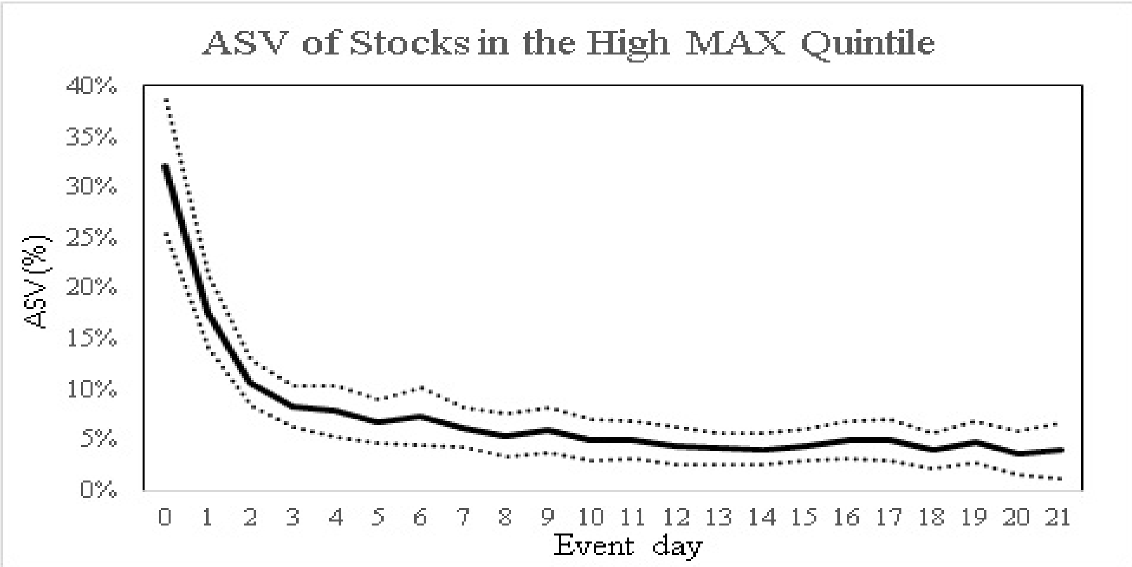
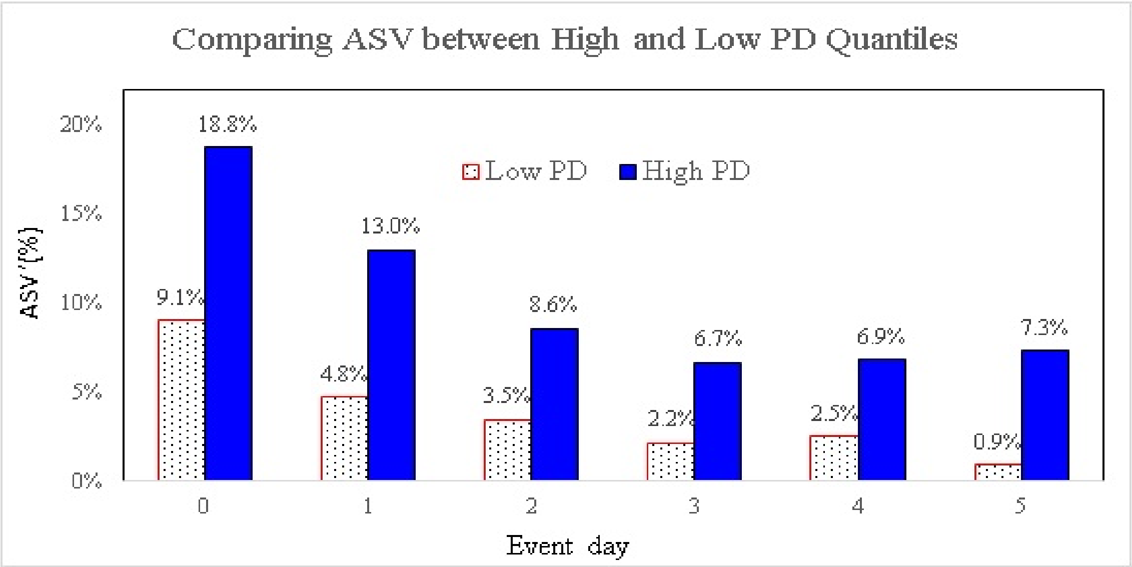


Figure 2. Abnormal Google search volume for the MAX event. This figure depicts the average daily abnormal Google search volume (ASV) on the MAX day (day 0) and the subsequent 21 trading days of stocks within the highest value-weighted MAX quintile portfolio. The dashed lines are the upper and lower 95% confidence bounds.



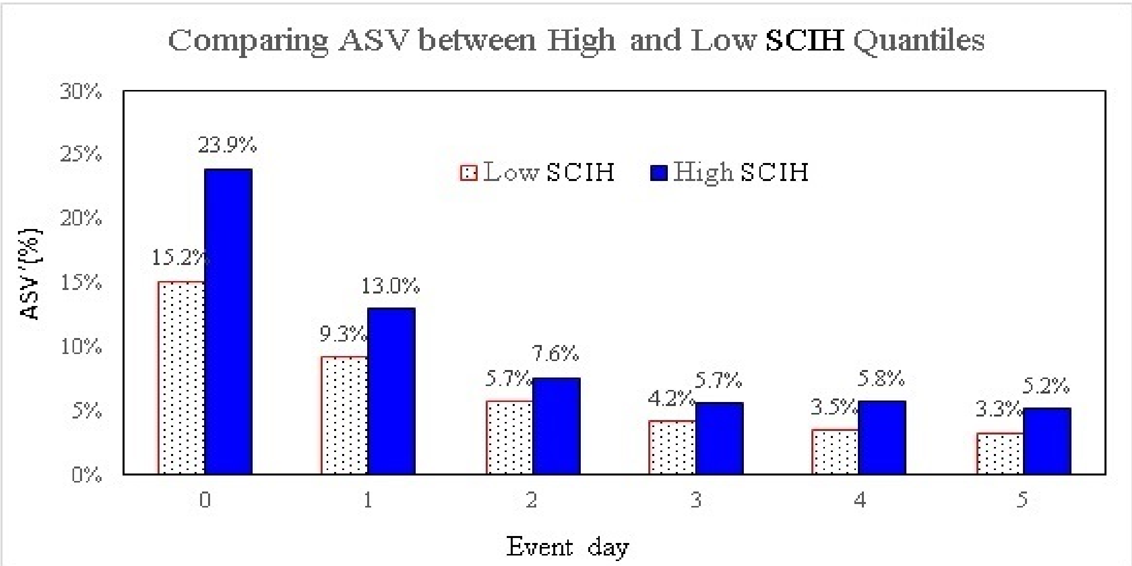


Figure 3. Comparing abnormal Google search volume between high and low social-interaction groups. The solid (dotted) bars depict the average daily abnormal Google search volume (ASV) on the MAX day (day 0) and the subsequent five trading days of stocks within the highest value-weighted MAX quintile portfolio and the high (low) social-interaction-based groups. The headquarters social interaction proxies are population density (PD) in the upper panel and Facebook social connectedness index (SCIH) in the lower panel.

Table 1
Descriptive statistics

In Panel A, the first three columns report the time series averages of the cross-sectional mean, median and standard deviation of each variable used in the paper, and the last two columns report the beginning and the ending months for each variable. Panel B reports the time series averages of the monthly cross-sectional correlations (multiplied by 100) of the variables. The lottery characteristics include the highest percentage daily return in a month (MAX), per-share stock price (PRC), idiosyncratic volatility measured in percentage terms (IVOL), and idiosyncratic skewness (ISKEW). The investor clientele proxy is the percentage of shares held by retail investors (RHLD). The attention proxies include the number of analyst covering (CVRG) and its natural logarithm (LNCVRG), the absolute value of earnings surprises (|SUE|), the number of trading days between the MAX day and the end of the portfolio formation month (DAYS), the recency of the lottery event (RECENCY), headquarters population density measured in thousand population per square mile (PD), and headquarters Facebook social connectedness index (SCIH). The set of stock return predictors include market beta (β^{MKT}), market capitalization measured in millions of dollars (SIZE), book-to-market (BM), momentum (MOM), illiquidity (ILLIQ), co-skewness (COSKEW), and standardized unexpected earnings (SUE). The proxies for a county's socioeconomic conditions are percent of Catholic population (CATH), percent of the population aged 25 and older who attained a bachelor's degree or higher (EDU), and median household income measured in thousands of dollars (MHI).

Variable	Mean	Median	Std. dev.	Begin	End
Lottery characteristics					
MAX	5.73	4.57	5.11	Jun, '63	Nov, '17
Investor clientele					
RHLD	57.48	56.59	25.97	Jan, '80	Sep, '17
Attention and social interaction proxies					
CVRG	7.34	4.87	6.92	Jan, '72	Nov, '17
LNCVRG	1.53	1.56	0.99	Jan, '72	Nov, '17
|SUE|	0.84	0.71	0.61	Jul, '74	Nov, '17
DAYS	10.01	10.12	6.16	Jun, '63	Nov, '17
RECENCY	0.19	0.10	0.23	Jun, '63	Nov, '17
PD	4.89	1.35	13.39	Jan, '76	Dec, '15
SCIH	0.37	0.24	0.42	Jun, '63	Nov, '17
Return predictors					
β^{MKT}	1.27	1.16	0.81	Jun, '63	Nov, '17
SIZE	1,967	278	8,243	Jun, '63	Nov, '17
BM	0.84	0.70	0.88	Jun, '63	Nov, '17
MOM	20.81	11.54	54.38	Jun, '63	Nov, '17
ILLIQ	1.81	0.20	11.73	Jun, '63	Nov, '17
COSKEW	-0.03	-0.04	0.23	Jun, '63	Nov, '17
SUE	-0.03	-0.02	1.03	Jul, '74	Nov, '17
IVOL	2.11	1.81	1.40	Jun, '63	Nov, '17
Socioeconomic variables					
CATH	25.02	22.61	14.59	Jan, '76	Dec, '15
EDU	28.19	26.87	8.64	Jan, '76	Dec, '15
MHI	40.72	38.39	9.54	Jan, '76	Dec, '15

Table 1 – continued

Variable		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)
(1)	MAX	100	-18.1	87.0	40.3	12.1	-15.5	-15.8	1.7	-2.1	-3.5	-1.2	2.2	24.2	-9.0	0.3	2.0	13.0	-0.8	1.8	1.2	5.6	7.6
(2)	PRC		100	-26.6	-1.5	-30.2	39.3	36.5	1.6	-2.0	-2.6	7.7	1.7	-14.9	34.7	-12.6	11.8	-17.6	3.7	4.6	1.7	6.0	-0.3
(3)	IVOL			100	15.8	18.3	-21.9	-22.5	2.5	0.3	-1.4	-2.5	2.1	28.2	-13.3	0.2	2.8	19.1	-1.8	-0.7	1.0	5.8	8.9
(4)	ISKEW				100	3.1	-4.2	-4.4	-0.7	-2.5	-4.2	0.4	0.7	4.0	-2.0	1.8	0.1	0.6	-1.3	4.2	0.8	1.1	1.2
(5)	RHLD					100	-42.6	-51.0	-1.1	2.7	3.9	-4.3	-8.0	-11.0	-14.4	14.2	3.0	19.7	-13.5	-1.1	-3.4	-11.7	-7.3
(6)	CVRG						100	89.4	1.4	-2.1	-3.2	8.6	7.0	-7.9	47.9	-14.2	-5.9	-16.9	16.4	0.3	-0.2	10.4	3.7
(7)	LNCVRG							100	1.7	-2.4	-3.4	6.5	6.5	-4.1	34.1	-17.9	-6.2	-23.7	15.8	-0.2	-0.7	10.1	3.7
(8)	|SUE|								100	0.0	-0.1	-0.5	-0.2	0.1	0.7	-2.5	-2.8	-0.9	1.0	-6.1	-0.4	0.0	0.2
(9)	NDAYS									100	71.7	-0.6	-0.6	-1.5	-1.1	0.9	0.0	2.3	-0.6	-0.1	-0.2	-0.9	-0.5
(10)	RECENCY										100	-0.6	-0.6	-2.4	-1.8	1.4	-0.5	2.0	-1.1	-0.3	-0.2	-1.0	-0.6
(11)	PD											100	23.5	2.3	7.9	1.2	-0.5	-1.5	2.1	0.1	10.1	45.4	-9.0
(12)	SCIH												100	6.6	2.8	-0.5	0.1	-3.4	0.5	-0.1	15.6	4.1	-10.1
(13)	BETA													100	-8.2	-7.2	4.9	-2.7	4.5	-0.6	3.5	12.3	13.1
(14)	SIZE														100	-7.2	-0.1	-6.4	5.6	1.0	2.2	7.1	2.7
(15)	BM															100	3.4	15.2	-0.5	4.1	-0.3	-6.4	-7.7
(16)	MOM																100	-5.5	-2.3	21.2	-0.1	1.0	1.1
(17)	ILLIQ																	100	-4.0	-0.2	-1.0	-5.2	-2.5
(18)	COSKEW																		100	-0.2	0.2	4.3	2.7
(19)	SUE																			100	-0.1	0.3	0.2
(20)	CATH																				100	17.0	33.7
(21)	EDU																					100	59.4
(22)	MHI																						100

Table 2
Bivariate portfolios of stocks sorted by retail holdings and MAX

For each month, all stocks in the sample are grouped into three portfolios with tercile breakpoints based on an ascending sort of retail holdings (RHLD). Stocks are then independently sorted into quintile portfolios based on an ascending sort of MAX. The intersections of the three RHLD-based groups and the five lottery-based groups generate a total of 15 value-weighted portfolios. This table reports the FF5 alphas of value-weighted average monthly excess returns (in percentages) for individual portfolios and the return difference between quintile 5 (High) and quintile 1 (Low) lottery-based portfolios within each RHLD-based group. Newey-West adjusted t-statistics are given in parentheses.

MAX	Low RHLD	Medium	High RHLD
Low	-0.01	0.15	0.06
	(-0.07)	(2.14)	(0.55)
2	-0.05	-0.03	0.10
	(-0.68)	(-0.45)	(0.69)
3	0.06	0.10	0.17
	(0.66)	(0.81)	(1.00)
4	0.02	-0.17	-0.09
	(0.17)	(-1.45)	(-0.64)
High	0.15	-0.36	-0.78
	(1.15)	(-2.52)	(-3.46)
High−Low	0.16	-0.51	-0.84
	(0.89)	(-3.01)	(-3.48)

Table 3
Retail portfolios sorted by investor attention and MAX

At the end of each month, stocks with the highest tercile of retail holdings (RHLD) are partitioned into three attention (ATNT) groups with tercile breakpoints and, independently, into quintile groups based on MAX. Each panel presents the FF5 alphas for the 15 portfolios from the intersections of the three attention-based groups and the five lottery-based groups. The last row reports the return difference between quintile 5 (High) and quintile 1 (Low) lottery-based portfolios within each attention-based group. The attention variables are analyst coverage (CVRG), absolute value of the standardized earnings surprise (|SUE|), and the recency of the lottery event (RECENCY). Newey-West adjusted t-statistics are given in parentheses.

	ATNT = CVRG			ATNT = |SUE|			ATNT = RECENCY		
MAX	Low	Medium	High	Low	Medium	High	Low	Medium	High
Low	0.10	0.18	0.06	0.15	0.15	0.28	0.15	0.15	0.00
	(0.96)	(1.64)	(0.50)	(0.99)	(1.13)	(2.00)	(1.09)	(1.11)	(-0.03)
2	0.19	0.07	-0.01	0.08	0.15	-0.26	-0.02	0.17	0.15
	(1.36)	(0.54)	(-0.09)	(0.42)	(0.92)	(-1.46)	(-0.12)	(1.08)	(0.90)
3	0.08	0.12	-0.01	0.20	-0.49	-0.02	0.19	0.13	0.10
	(0.52)	(0.81)	(-0.03)	(0.95)	(-3.06)	(-0.10)	(0.74)	(0.77)	(0.44)
4	-0.33	-0.35	-0.41	-0.36	0.07	-0.47	-0.07	-0.10	-0.16
	(-2.35)	(-1.95)	(-1.09)	(-1.84)	(0.26)	(-2.07)	(-0.31)	(-0.54)	(-0.71)
High	-0.65	-0.57	-1.26	-0.63	-0.43	-0.80	-0.46	-0.72	-1.47
	(-2.82)	(-2.29)	(-2.24)	(-2.80)	(-1.90)	(-2.92)	(-1.61)	(-2.96)	(-4.53)
High−Low	-0.75	-0.75	-1.33	-0.78	-0.58	-1.09	-0.61	-0.87	-1.47
	(-2.76)	(-2.78)	(-2.24)	(-2.52)	(-2.26)	(-3.61)	(-1.91)	(-3.14)	(-4.01)

Table 4
Retail portfolios sorted by social interaction and MAX

At the end of each month, stocks with the highest tercile of retail holdings (RHLD) are partitioned based on a social interaction variable (SOCIAL) into three groups with tercile breakpoints and independently into quintile groups based on MAX. Each panel presents the FF5 alphas for the 15 portfolios from the intersections of the three SOCIAL-based groups and the five lottery-based groups. The last row reports the return difference between quintile 5 (High) and quintile 1 (Low) lottery-based portfolios within each attention-based group. The social interaction variables are headquarters population density (PD) and Facebook social connectedness index (SCIH). Newey-West adjusted t-statistics are given in parentheses.

	SOCIAL = PD			SOCIAL = SCIH		
MAX	Low	Medium	High	Low	Medium	High
Low	0.17	0.14	0.07	0.27	0.22	-0.02
	(1.26)	(1.09)	(0.50)	(1.93)	(1.67)	(-0.14)
2	0.05	0.31	-0.18	0.18	0.07	-0.24
	(0.32)	(1.51)	(-0.93)	(1.16)	(0.38)	(-1.27)
3	0.14	0.11	0.29	0.03	0.02	0.26
	(0.82)	(0.45)	(1.16)	(0.20)	(0.07)	(1.08)
4	-0.37	0.12	-0.42	-0.43	0.01	-0.27
	(-1.65)	(0.56)	(-1.79)	(-2.34)	(0.05)	(-1.06)
High	-0.49	-0.73	-1.29	-0.64	-0.70	-1.43
	(-1.88)	(-2.13)	(-4.71)	(-3.01)	(-2.22)	(-4.39)
High−Low	-0.66	-0.87	-1.35	-0.91	-0.91	-1.41
	(-2.55)	(-2.35)	(-4.73)	(-3.33)	(-2.64)	(-4.09)

Table 5
Fama-MacBeth regressions: Investor attention and lottery stock returns

This table reports the time series averages of the slope coefficients obtained from regressing monthly excess returns (in percentages) on a lagged lottery variable measured by MAX and its interaction with a lagged attention proxy (ATNT) and a retail holding dummy. D^H, D^M, and D^L are the retail holding dummy variables, set equal to one if a stock's RHLD is in the top tercile, medium tercile, and bottom tercile of RHLD-based groups, respectively, and zero otherwise. The control variables include market beta (β^{MKT}), natural log of market capitalization (SIZE), natural logarithm of book-to-market (BM), momentum (MOM), illiquidity (ILLIQ), co-skewness (COSKEW), standardized unexpected earnings (SUE), and idiosyncratic volatility (IVOL). The proxies for investor attention are the natural logarithm of the number of analysts covering a stock (LNCVRG), the absolute value of the standardized quarterly unexpected earnings (|SUE|), and the recency of the lottery event (RECENCY). The last two rows report the average number of monthly observations used in the Fama-MacBeth regressions and the average adjusted R-squared from the Fama-MacBeth regressions, respectively. Newey-West adjusted t-statistics are given in parentheses.

			ATNT = LNCVRG		ATNT = |SUE|		ATNT = RECENCY	
Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
MAX×ATNT×D^H			-0.0400	-0.0387	-0.0387	-0.0418	-0.1595	-0.2077
			(-3.52)	(-4.17)	(-4.23)	(-4.99)	(-5.01)	(-6.49)
MAX×ATNT×D^M			-0.0115	-0.0163	-0.0028	-0.0132	-0.0685	-0.1382
			(-1.19)	(-2.38)	(-0.25)	(-1.53)	(-1.84)	(-4.50)
MAX×ATNT×D^L			0.0150	0.0052	0.0141	0.0016	0.0072	-0.0646
			(1.72)	(0.76)	(1.26)	(0.18)	(0.20)	(-2.05)
MAX	-0.0709	-0.0547	-0.0636	-0.0267	-0.0488	-0.0292	-0.0502	-0.0190
	(-5.88)	(-4.98)	(-5.75)	(-1.94)	(-5.15)	(-2.64)	(-4.58)	(-1.83)
ATNT			-0.0987	0.1484	-0.0131	0.0989	-0.3756	-0.1137
			(-1.57)	(3.02)	(-0.28)	(2.79)	(-2.14)	(-0.90)
RHLD			-0.0099	-0.0071	-0.0051	-0.0073	-0.0049	-0.0048
			(-5.56)	(-4.24)	(-2.76)	(-4.60)	(-2.49)	(-2.89)
β^{MKT}		0.1280		0.0254		0.0478		0.0542
		(1.34)		(0.27)		(0.50)		(0.57)
SIZE		-0.1268		-0.1923		-0.1347		-0.1450
		(-4.71)		(-4.82)		(-4.45)		(-4.77)
BM		0.0959		0.0507		0.0809		0.0802
		(1.60)		(0.79)		(1.37)		(1.37)
MOM		0.0045		0.0051		0.0048		0.0048
		(3.69)		(3.53)		(3.64)		(3.63)
ILLIQ		-0.0190		-0.0518		-0.0071		-0.0058
		(-2.60)		(-3.16)		(-1.02)		(-0.88)
COSKEW		-0.2224		-0.1435		-0.1661		-0.1697
		(-1.99)		(-1.23)		(-1.43)		(-1.47)
SUE		0.3173		0.2391		0.2862		0.2827
		(14.38)		(11.43)		(13.86)		(13.73)
IVOL		-0.1037		-0.1384		-0.1131		-0.1292
		(-2.32)		(-3.04)		(-2.59)		(-2.93)
N	3,452	2,334	2,982	2,112	2,535	2,462	3,848	2,462
Adj. R^2	0.01	0.05	0.15	0.19	0.14	0.17	0.13	0.17

Table 6

Fama-MacBeth regressions: Social interactions and lottery stock returns

This table reports the time series averages of the slope coefficients obtained from regressing monthly excess returns (in percentages) on lagged MAX and its interaction with a lagged social interaction proxy (SOCIAL) and a retail holding dummy. D^H, D^M, and D^L are the retail holding dummy variables, set equal to one if a stock's RHLD is in the top tercile, medium tercile, and bottom tercile of RHLD-based groups, respectively, and zero otherwise. The control variables include market beta (β^{MKT}), the natural log of market capitalization (SIZE), the natural logarithm of book-to-market (BM), momentum (MOM), illiquidity (ILLIQ), co-skewness (COSKEW), standardized unexpected earnings (SUE), and idiosyncratic volatility (IVOL). The proxies for a firm's headquarters social interactions are population density (PD) and Facebook social connectivity (SCIH). The slope coefficients of the PD-related interaction terms are multiplied by 100. The last two rows report the average number of monthly observations used in the Fama-MacBeth regressions and the average adjusted R-squared from the Fama-MacBeth regressions, respectively. Newey-West adjusted t-statistics are given in parentheses.

	SOCIAL = PD		SOCIAL = SCIH	
Variable	(1)	(2)	(3)	(4)
MAX×SOCIAL×D^H	-0.1395	-0.1656	-0.0679	-0.0554
	(-2.36)	(-2.32)	(-4.42)	(-3.50)
MAX×SOCIAL×D^M	-0.0470	-0.1190	-0.0149	-0.0284
	(-0.71)	(-2.13)	(-0.87)	(-1.88)
MAX×SOCIAL×D^L	-0.0222	-0.0404	0.0103	0.0006
	(-0.39)	(-0.75)	(0.66)	(0.04)
MAX	-0.0670	-0.0424	-0.0573	-0.0385
	(-5.24)	(-3.61)	(-5.15)	(-3.42)
SOCIAL	0.0007	0.0038	0.0166	0.0775
	(0.33)	(2.04)	(0.23)	(1.21)
RHLD	-0.0045	-0.0049	-0.0046	-0.0046
	(-2.12)	(-2.77)	(-2.22)	(-2.75)
β^{MKT}		0.0337		0.0456
		(0.33)		(0.47)
SIZE		-0.1455		-0.1394
		(-4.59)		(-4.59)
BM		0.0840		0.0836
		(1.38)		(1.42)
MOM		0.0055		0.0049
		(4.25)		(3.68)
ILLIQ		-0.0073		-0.0070
		(-1.02)		(-1.03)
COSKEW		-0.1601		-0.1540
		(-1.32)		(-1.32)
SUE		0.2970		0.2844
		(14.14)		(13.69)
IVOL		-0.1322		-0.1101
		(-2.99)		(-2.54)
N	3,642	2,380	3,656	2,404
Adj. R^2	0.13	0.17	0.13	0.17

Table 7
Social interactions and lottery stock returns, controlling for investor preferences

This table reports the time series averages of the slope coefficients obtained from Fama-MacBeth regressions of monthly excess returns (in percentages) on a lagged MAX and its interaction with a lagged social interaction proxy (SOCIAL) and a retail holding dummy. D^H, D^M, and D^L are the retail holding dummy variables, set equal to one if a stock's RHLD is in the top tercile, medium tercile, and bottom tercile of RHLD-based groups, respectively, and zero otherwise. The control variables include market beta (β^{MKT}), the natural log of market capitalization (SIZE), the natural logarithm of book-to-market (BM), momentum (MOM), illiquidity (ILLIQ), co-skewness (COSKEW), standardized unexpected earnings (SUE), and idiosyncratic volatility (IVOL). The proxies for a firm's headquarters social interactions are population density (PD) in Column 1 and Facebook social connectivity (SCIH) in Column 2. The proxies for investor preferences are Catholic population (CATH) measured in percentage terms, percent of the population aged 25 and older who attained a bachelor's degree or higher (EDU), and median household income measured in thousands of dollars (MHI). The slope coefficients of the PD-related interaction terms are multiplied by 100. The last two rows report the average number of monthly observations used in the Fama-MacBeth regressions and the average adjusted R-squared from the Fama-MacBeth regressions, respectively. Newey-West adjusted t-statistics are given in parentheses.

	SOCIAL = PD	SOCIAL = SCIH
Variable	(1)	(2)
MAX×SOCIAL×D^H	-0.1617	-0.0569
	(-2.27)	(-3.53)
MAX×SOCIAL×D^M	-0.1175	-0.0312
	(-2.11)	(-2.02)
MAX×SOCIAL×D^L	-0.0369	-0.0012
	(-0.70)	(-0.08)
CATH	-0.0006	-0.0007
	(-0.46)	(-0.63)
EDU	0.0049	0.0036
	(1.60)	(1.77)
MHI	0.0012	0.0005
	(0.30)	(0.15)
Controls	MAX, SOCIAL, RHLD, β^{MKT},SIZE, BM MOM, ILLIQ, COSKEW, SUE, IVOL	
N	2,370	2,353
Adj. R^2	0.18	0.17

Table 8
Return responses to earnings announcements

For each retail-dominated lottery stock(i.e., stocks in the highest MAX quintile group and the top tercile RHLD group in month t) with an earnings announcement in month $t+1$, we calculate the cumulative market-adjusted return (CAR) for the three days surrounding the earnings announcement, and the cumulative market-adjusted return after excluding the three-day earnings announcement return (RET−CAR) for the month. For each month t, we partition lottery stocks into three portfolios with tercile breakpoints based on an ascending sort of a proxy for investor attention or social interaction. The attention proxies are analyst coverage (CVRG) and the recency of the lottery event (RECENCY). The proxies for a firm's headquarters social interactions are population density (PD) and Facebook social connectivity (SCIH). Panel A reports the average CAR for each value-weighted portfolio. Panel B reports the average RET−CAR and the FF5 alpha for each value-weighted portfolio. The sample period is October 1971 through December 2017. Newey-West adjusted t-statistics are given in parentheses.

Panel A. Abnormal earnings-announcement returns

	CVRG			RECENCY			PD			SCIH		
	Low	Medium	High	Low	Medium	High	Low	Medium	High	Low	Medium	High
CAR	-0.61	-0.47	-1.54	-0.01	-0.40	-0.64	0.05	-0.20	-0.91	0.03	-0.46	-0.68
	(-3.57)	(-2.39)	(-3.16)	(-0.04)	(-2.40)	(-2.95)	(0.25)	(-0.95)	(-3.82)	(0.12)	(-2.19)	(-2.93)

Panel B. Abnormal monthly returns after excluding earnings announcement returns

	CVRG			RECENCY			PD			SCIH		
	Low	Medium	High	Low	Medium	High	Low	Medium	High	Low	Medium	High
RET−CAR	0.36	0.72	0.18	0.38	-1.00	0.16	-0.51	0.20	0.52	0.22	0.19	-0.28
	(1.65)	(2.65)	(0.48)	(1.37)	(-3.33)	(0.72)	(-1.43)	(0.93)	(1.87)	(0.61)	(0.66)	(-0.89)
FF5 alpha	0.25	0.63	0.18	0.33	-0.94	0.16	-0.55	0.11	0.36	0.21	0.10	-0.48
	(1.10)	(1.93)	(0.41)	(1.13)	(-2.84)	(0.74)	(-1.55)	(0.53)	(1.29)	(0.53)	(0.37)	(-1.27)

Table 9
Realizations of earnings surprises

For each retail-dominated lottery stock(i.e., stocks in the highest MAX quintile group and the top tercile RHLD group in month t) with an earnings announcement in month $t+1$, we calculate its standardized earnings surprise (SUE) based on the seasonal-random-walk model. For each month t, we partition lottery stocks into three portfolios with tercile breakpoints based on an ascending sort of a proxy for investor attention or social interactions. The attention proxies are analyst coverage (CVRG) and the recency of the lottery event (RECENCY). The proxies for a firm's headquarters social interactions are population density (PD) and Facebook social connectivity (SCIH). This table reports the average SUE for each value-weighted portfolio. The sample period is August 1974 through December 2017. Newey-West adjusted t-statistics are given in parentheses.

	CVRG			RECENCY			PD			SCIH		
	Low	Medium	High	Low	Medium	High	Low	Medium	High	Low	Medium	High
SUE	-0.001	-0.007	-0.075	0.004	-0.014	-0.029	-0.002	-0.008	-0.031	-0.003	-0.018	-0.005
	(-0.11)	(-0.44)	(-1.83)	(0.22)	(-0.79)	(-1.65)	(-0.08)	(-0.45)	(-1.57)	(-0.18)	(-0.90)	(-0.24)

Table 10
Retail order imbalances following MAX

In Panel A, for each month, we calculate the average retail order imbalances on the day following MAX across lottery stocks (i.e., stocks in the highest MAX quintile group) and the average retail order imbalances on the day following earnings announcements across stocks making earnings announcements in the month. Panel A reports the time series averages of the monthly differences in the volume-based (OIBVOL) and trade-based (OIBTRD) retail order imbalances between the MAX event and the earnings announcement event. In Panel B, for each month, we partition lottery stocks into three portfolios with tercile breakpoints based on an ascending sort of a proxy for investor attention or social interactions. The attention proxies are analyst coverage (CVRG), absolute value of the standardized earnings surprise (|SUE|), and the recency of the lottery event (RECENCY). The proxies for a firm's headquarters social interactions are population density (PD) and Facebook social connectivity (SCIH). We calculate the average retail order imbalance on the day following MAX across stocks in the same portfolio. Panel B reports the time series averages of the monthly average retail order imbalances for each portfolio. The retail order imbalances are measured based on share volume (OIBVOL) and the number of trades (OIBTRD). The sample period is January 2010 through December 2017. Newey-West adjusted t-statistics are given in parentheses.

Panel A. Differences in order imbalances between lottery event and earnings announcement event

OIBVOL	OIBTRD
1.72%	2.10%
(6.43)	(6.87)

Panel B. Retail order imbalances controlling for attention/social interaction

	OIBVOL					OIBTRD				
	CVRG	|SUE|	RECENCY	PD	SCIH	CVRG	|SUE|	RECENCY	PD	SCIH
Low	0.25%	-0.08%	0.09%	0.37%	0.36%	0.35%	-0.25%	0.47%	0.38%	0.26%
	(0.84)	(-0.29)	(0.22)	(1.05)	(0.81)	(1.26)	(-1.01)	(1.34)	(0.99)	(0.65)
Medium	0.48%	0.24%	0.45%	0.18%	-0.18%	0.54%	0.30%	0.39%	0.12%	-0.06%
	(1.41)	(0.98)	(2.12)	(0.54)	(-0.66)	(1.60)	(0.97)	(1.57)	(0.47)	(-0.21)
High	0.47%	1.32%	0.53%	0.85%	1.33%	0.69%	1.00%	0.67%	1.01%	1.28%
	(0.99)	(2.81)	(1.26)	(2.21)	(4.16)	(1.66)	(2.82)	(1.92)	(2.94)	(4.45)

Table 11
Accounting for arbitrage costs

This table reports the time series averages of the slope coefficients obtained from Fama-MacBeth regressions of monthly excess returns (in percentages) on lagged MAX and its interaction with a lagged proxy for investor attention (ATNT) or social interaction proxy (SOCIAL) and a retail ownership dummy after controlling for the arbitrage cost score (COST). D^H, D^M, and D^L are the retail ownership dummy variables, set equal to one if a stock's RHLD is in the top tercile, medium tercile, and bottom tercile RHLD-based groups, respectively, and zero otherwise. The proxies for investor attention are the natural logarithm of the number of analysts covering a stock (LNCVRG), the absolute value of the standardized quarterly unexpected earnings (|SUE|), and the recency of the lottery event (RECENCY). The proxies for a firm's headquarters' social interactions are population density (PD) and Facebook social connectivity (SCIH). The control variables include market beta (β^{MKT}), natural log of market capitalization (SIZE), natural logarithm of book-to-market (BM), momentum (MOM), illiquidity (ILLIQ), co-skewness (COSKEW), standardized unexpected earnings (SUE), and idiosyncratic volatility (IVOL). The slope coefficients of the PD-related interaction terms are multiplied by 100. The last two rows report the average number of monthly observations used in the Fama-MacBeth regressions and the average adjusted R-squared from the Fama-MacBeth regressions, respectively. Newey-West adjusted t-statistics are given in parentheses.

	Investor attention proxies			Social interaction proxies	
	LNCVRG	|SUE|	RECENCY	PD	SCIH
Variable	(1)	(2)	(3)	(4)	(5)
MAX×ATNT/SOCIAL×D^H	-0.0397	-0.0415	-0.2067	-0.1631	-0.0541
	(-4.35)	(-4.97)	(-6.53)	(-2.27)	(-3.41)
MAX×ATNT/SOCIAL×D^M	-0.0175	-0.0131	-0.1408	-0.1105	-0.0264
	(-2.57)	(-1.54)	(-4.56)	(-1.99)	(-1.76)
MAX×ATNT/SOCIAL×D^L	0.0030	0.0015	-0.0669	-0.0425	-0.0003
	(0.45)	(0.17)	(-2.09)	(-0.80)	(-0.02)
COST	0.0487	-0.0146	0.0059	-0.0051	-0.0023
	(1.05)	(-0.39)	(0.16)	(-0.13)	(-0.06)
Controls	MAX, ATNT/SOCIAL, RHLD, β^{MKT},SIZE BM, MOM, ILLIQ, COSKEW, SUE, IVOL				
N	2,112	2,462	2,462	2,380	2,404
Adj. R^2	0.19	0.17	0.17	0.17	0.17

Table 12
Accounting for information supply

This table reports the time series averages of the slope coefficients obtained from regressing monthly excess returns (in percentages) on lagged MAX and its interaction with a lagged proxy for investor attention (ATNT) or social interaction (SOCIAL) and a retail holding dummy. D^H, D^M, and D^L are the retail holding dummy variables, set equal to one if a stock's RHLD is in the top tercile, medium tercile, and bottom tercile RHLD-based groups, respectively, and zero otherwise. The control variables include market beta (β^{MKT}), the natural log of market capitalization (SIZE), the natural logarithm of book-to-market (BM), momentum (MOM), illiquidity (ILLIQ), co-skewness (COSKEW), standardized unexpected earnings (SUE), idiosyncratic volatility (IVOL), and information supply measured by the number of relevant news reports (NEWS) from credible sources as defined by Ravenpack. The Ravenpack news data is available for the period of January 2000 through December 2017. The proxies for a firm's headquarters social interactions are population density (PD) and Facebook social connectivity (SCIH). The slope coefficients of the PD-related interaction terms are multiplied by 100. The last two rows report the average number of monthly observations used in the Fama-MacBeth regressions and the average adjusted R-squared from the Fama-MacBeth regressions, respectively. Newey-West adjusted t-statistics are given in parentheses.

	Attention proxies			Social interaction proxies	
	LNCVRG	|SUE|	RECENCY	PD	SCIH
Variable	(1)	(2)	(3)	(4)	(5)
MAX×ATNT/SOCIAL×D^H	-0.0633	-0.0350	-0.1474	-0.2346	-0.0751
	(-4.10)	(-2.43)	(-3.04)	(-2.72)	(-2.54)
MAX×ATNT/SOCIAL×D^M	-0.0281	-0.0053	-0.0526	-0.2433	-0.0457
	(-2.63)	(-0.41)	(-1.06)	(-3.09)	(-1.74)
MAX×ATNT/SOCIAL×D^L	-0.0115	-0.0018	0.0497	-0.1126	-0.0130
	(-1.11)	(-0.15)	(1.13)	(-1.46)	(-0.51)
NEWS	0.095	0.087	0.086	0.093	0.084
	(2.75)	(2.60)	(2.58)	(2.65)	(2.50)
Controls	MAX, ATNT/SOCIAL, RHLD, β^{MKT},SIZE BM, MOM, ILLIQ, COSKEW, SUE, IVOL				
N	2,090	2,339	2,339	2,292	2,305
Adj. R^2	0.19	0.18	0.18	0.19	0.18

Table 13
Accounting for microstructure effects

Panels A and B report the time series averages of the slope coefficients obtained from monthly ordinary Fama-MacBeth regressions using a sample that excludes microcap stocks, and monthly weighted least squares (WLS) regressions using the full sample. In both tests, monthly excess returns (in percentages) are regressed on lagged MAX and its interaction with a lagged investor attention proxy (ATNT) or social interaction proxy (SOCIAL) and a retail holding dummy. D^H, D^M, and D^L are the retail holding dummy variables, set equal to one if a stock's RHLD is in the top tercile, medium tercile, and bottom tercile RHLD-based groups, respectively, and zero otherwise. The proxies for investor attention are the natural logarithm of the number of analysts covering a stock (LNCVRG), the absolute value of the standardized quarterly unexpected earnings (|SUE|), and the recency of the lottery event (RECENCY). The proxies for a firm's headquarters' social interactions are population density (PD) and Facebook social connectivity (SCIH). The control variables include market beta (β^{MKT}), the natural log of market capitalization (SIZE), the natural logarithm of book-to-market (BM), momentum (MOM), illiquidity (ILLIQ), co-skewness (COSKEW), standardized unexpected earnings (SUE), and idiosyncratic volatility (IVOL). The slope coefficients of the PD-related interaction terms are multiplied by 100. The last two rows report the average number of monthly observations used in the Fama-MacBeth regressions and the average adjusted R-squared from the Fama-MacBeth regressions, respectively. Newey-West adjusted t-statistics are given in parentheses.

Panel A. Fama-MacBeth regressions excluding microcaps

	Attention proxies			Social interaction proxies	
	LNCVRG	|SUE|	RECENCY	PD	SCIH
variable	(1)	(2)	(3)	(4)	(5)
MAX×ATNT/SOCIAL×D^H	-0.0347	-0.0429	-0.2088	-0.1451	-0.0664
	(-3.36)	(-3.17)	(-3.81)	(-1.69)	(-2.55)
MAX×ATNT/SOCIAL×D^M	-0.0137	-0.0203	-0.2042	-0.0569	-0.0240
	(-1.77)	(-1.78)	(-4.06)	(-0.81)	(-1.19)
MAX×ATNT/SOCIAL×D^L	-0.0012	-0.0006	-0.1233	-0.0522	-0.0210
	(-0.15)	(-0.05)	(-2.81)	(-0.82)	(-1.09)
Controls	MAX, ATNT/SOCIAL, RHLD, β^{MKT},SIZE BM, MOM, ILLIQ, COSKEW, SUE, IVOL				
N	1,438	1,480	1,480	1,416	1,441
Adj. R^2	0.22	0.21	0.21	0.22	0.22

Table 13 – continued

Panel B. WLS regressions

	Attention proxies			Social interaction proxies	
	LNCVRG	|SUE|	RECENCY	PD	SCIH
variable	(1)	(2)	(3)	(4)	(5)
MAX×ATNT/SOCIAL×D^H	-0.0351	-0.0433	-0.1953	-0.1897	-0.0598
	(-3.88)	(-4.66)	(-5.74)	(-2.59)	(-3.74)
MAX×ATNT/SOCIAL×D^M	-0.015	-0.0137	-0.1222	-0.1247	-0.0257
	(-2.15)	(-1.47)	(-3.70)	(-2.13)	(-1.62)
MAX×ATNT/SOCIAL×D^L	0.0062	0.0001	-0.0509	-0.0499	0.0015
	(0.89)	(0.01)	(-1.51)	(-0.89)	(0.09)
Controls	MAX, ATNT/SOCIAL, RHLD, β^{MKT},SIZE BM, MOM, ILLIQ, COSKEW, SUE, IVOL				
N	2,112	2,462	2,462	2,380	2,404
Adj. R^2	0.19	0.17	0.17	0.17	0.17

Online Appendix

To save space in the paper, we present some of our findings in an online appendix.

- Table A1 reports the results from univariate portfolio sorts based on MAX.
- Table A2 reports the average value of MAX for each portfolio sorted by retail holdings and MAX.
- Table A3 presents the results from trivariate sorts by retail holdings, an attention proxy, and MAX.
- Table A4 reports the results from bivariate portfolio sorts based on the orthogonal component of analyst coverage and MAX.
- Table A5 presents the results from trivariate sorts by retail holdings, a social interaction proxy, and MAX.
- Table A6 presents the results from Fama-MacBeth regressions using a retail ownership measure that is orthogonlized to firm size.
- Table A7 reports the time series averages of the slope coefficients obtained from regressing monthly excess returns for retail stocks (i.e., stocks in the top tercile retail ownership group) on a lagged MAX and its interaction with a lagged attention proxy (ATNT) or a social interaction proxy.
- Table A8 compares the summary statistics for daily retail order imbalances between our sample and the sample of Boehmer et al. (2021).

Table A1
Univariate portfolios of stocks sorted by lottery proxies

For each month, decile portfolios are formed by sorting individual stocks based on our lottery proxy, the highest daily return in the previous month (MAX). The first row reports the value-weighted average monthly excess return (RET−RF) in percentage terms for each decile and the value-weighted average return difference between decile 10 (High) and decile 1 (Low); the last two rows present the corresponding alphas relative to the Fama-French-Carhart-Pastor-Stambaugh model (FFCPS) and the Fama-French five-factor model (FF5), respectively. Newey-West adjusted t-statistics are given in parentheses. The sample period is from July 1963 to December 2017.

MAX	Low	2	3	4	5	6	7	8	9	High	High−Low
RET−RF	0.61	0.57	0.59	0.50	0.63	0.64	0.58	0.47	0.46	-0.06	-0.66
	(4.34)	(3.62)	(3.48)	(2.70)	(3.16)	(3.01)	(2.40)	(1.72)	(1.48)	(-0.17)	(-2.41)
FFCPS	0.12	0.04	0.07	-0.07	0.06	0.05	-0.08	-0.15	-0.25	-0.78	-0.91
	(1.74)	(0.59)	(1.19)	(-1.18)	(0.74)	(0.71)	(-1.02)	(-1.34)	(-2.33)	(-4.81)	(-4.63)
FF5	0.07	-0.05	-0.03	-0.10	0.05	0.07	-0.03	-0.04	-0.04	-0.43	-0.50
	(1.14)	(-0.90)	(-0.45)	(-2.11)	(0.67)	(1.00)	(-0.40)	(-0.52)	(-0.40)	(-3.70)	(-3.60)

Table A2
Average MAX values for portfolios sorted by retail holdings and MAX

For each month, all stocks in the sample are grouped into three portfolios with tercile breakpoints based on an ascending sort of retail holdings (RHLD). Stocks are then independently sorted into quintile portfolios based on an ascending order of our lottery proxy, MAX. The intersections of the three RHLD-based groups and the five MAX-based groups generate a total of 15 value-weighted portfolios. This table reports the value-weighted averages of MAX, and the last row reports the the difference in MAX between quintile 5 (High) and quintile 1 (Low) portfolios within each retail holding group.

MAX	Low RHLD	Medium	High RHLD
Low	2.16	2.04	1.85
2	3.38	3.36	3.38
3	4.71	4.72	4.76
4	6.58	6.63	6.70
High	11.57	12.03	13.46
High−Low	9.41	9.99	11.61

Table A3
Portfolios sorted by retail holdings, investor attention, and MAX

At the end of each month, stocks in the sample are divided into three groups with tercile breakpoints (30%−40%−30%) based on an ascending sort of retail holdings (RHLD). Stocks are also independently partitioned into three groups with tercile breakpoints based on an ascending sort of an attention proxy (ATNT) and quintiles based on an ascending sort of MAX. The three-way portfolio sorts generate a total of 45 value-weighted portfolios. Each panel of this table presents the FF5 alphas for each of the 45 portfolios. The last row reports the return difference between quintile 5 (High) and quintile 1 (Low) MAX-based portfolios within each attention-based group. The attention variables are the analyst coverage (CVRG) in Panel A, the absolute value of the standardized earnings surprise ($|SUE|$) in Panel B, and the inverse of one plus the number of trading days between the MAX day and the last trading day in the portfolio formation month (RECENCY) in Panel C. Newey-West adjusted t-statistics are given in parentheses.

Panel A. CVRG as the attention proxy

	Low RHLD			Medium RHLD			High RHLD		
MAX	Low CVRG	Medium	High CVRG	Low CVRG	Medium	High CVRG	Low CVRG	Medium	High CVRG
Low	0.23	0.11	0.02	0.00	0.09	0.19	0.10	0.18	0.06
	(1.72)	(1.18)	(0.21)	(-0.04)	(0.98)	(2.43)	(0.96)	(1.64)	(0.50)
2	-0.09	-0.07	-0.04	0.10	-0.14	-0.08	0.19	0.07	-0.01
	(-0.57)	(-0.79)	(-0.48)	(1.00)	(-1.80)	(-0.95)	(1.36)	(0.54)	(-0.09)
3	0.21	0.00	0.05	0.15	-0.17	0.22	0.08	0.12	-0.01
	(1.32)	(0.00)	(0.43)	(1.01)	(-1.65)	(1.56)	(0.52)	(0.81)	(-0.03)
4	-0.16	-0.09	0.08	0.07	-0.19	-0.04	-0.33	-0.35	-0.41
	(-0.94)	(-0.82)	(0.66)	(0.48)	(-1.86)	(-0.31)	(-2.35)	(-1.95)	(-1.09)
High	-0.48	0.23	0.16	-0.17	-0.22	-0.05	-0.65	-0.57	-1.26
	(-1.67)	(1.43)	(0.97)	(-1.08)	(-1.52)	(-0.25)	(-2.82)	(-2.29)	(-2.24)
High−Low	-0.71	0.12	0.14	-0.17	-0.31	-0.24	-0.75	-0.75	-1.33
	(-2.54)	(0.55)	(0.70)	(-0.89)	(-1.82)	(-1.03)	(-2.76)	(-2.78)	(-2.24)

Table A3 – continued

Panel B. $|SUE|$ as the attention proxy

	Low RHLD			Medium RHLD			High RHLD		
MAX	Low $\lvert SUE \rvert$	Medium	High $\lvert SUE \rvert$	Low $\lvert SUE \rvert$	Medium	High $\lvert SUE \rvert$	Low $\lvert SUE \rvert$	Medium	High $\lvert SUE \rvert$
Low	-0.04	0.06	0.00	0.16	0.08	0.11	0.15	0.15	0.28
	(-0.28)	(0.64)	(-0.01)	(1.57)	(0.87)	(0.97)	(0.99)	(1.13)	(2.00)
2	0.02	-0.14	0.05	0.08	-0.11	-0.09	0.08	0.15	-0.26
	(0.21)	(-1.45)	(0.44)	(0.76)	(-0.85)	(-0.71)	(0.42)	(0.92)	(-1.46)
3	-0.11	0.05	-0.03	-0.27	-0.01	0.07	0.20	-0.49	-0.02
	(-0.86)	(0.49)	(-0.25)	(-1.87)	(-0.03)	(0.49)	(0.95)	(-3.06)	(-0.10)
4	0.08	-0.18	0.08	-0.18	0.08	-0.06	-0.36	0.07	-0.47
	(0.61)	(-1.23)	(0.54)	(-0.88)	(0.50)	(-0.33)	(-1.84)	(0.26)	(-2.07)
High	-0.05	-0.02	0.13	0.01	-0.33	-0.08	-0.63	-0.43	-0.80
	(-0.30)	(-0.13)	(0.64)	(0.03)	(-1.55)	(-0.37)	(-2.80)	(-1.90)	(-2.92)
High−Low	-0.02	-0.08	0.13	-0.15	-0.40	-0.20	-0.78	-0.58	-1.09
	(-0.07)	(-0.40)	(0.54)	(-0.52)	(-1.74)	(-0.80)	(-2.52)	(-2.26)	(-3.61)

Panel C. RECENCY as the attention proxy

	Low RHLD			Medium RHLD			High RHLD		
MAX	Low RECENCY	Medium	High RECENCY	Low RECENCY	Medium	High RECENCY	Low RECENCY	Medium	High RECENCY
Low	-0.07	0.05	-0.08	0.04	0.28	0.11	0.15	0.15	0.00
	(-0.66)	(0.45)	(-0.79)	(0.44)	(3.02)	(0.98)	(1.09)	(1.11)	(-0.03)
2	0.07	-0.02	-0.32	0.07	-0.05	-0.10	-0.02	0.17	0.15
	(0.80)	(-0.18)	(-2.48)	(0.56)	(-0.34)	(-0.65)	(-0.12)	(1.08)	(0.90)
3	-0.07	0.15	-0.24	0.12	0.23	-0.20	0.19	0.13	0.10
	(-0.65)	(1.26)	(-1.51)	(0.66)	(1.23)	(-0.87)	(0.74)	(0.77)	(0.44)
4	0.08	0.24	-0.32	0.09	-0.19	-0.78	-0.07	-0.10	-0.16
	(0.61)	(1.49)	(-1.87)	(0.58)	(-1.09)	(-3.34)	(-0.31)	(-0.54)	(-0.71)
High	0.29	0.27	-0.42	-0.13	-0.19	-0.89	-0.46	-0.72	-1.47
	(1.48)	(1.48)	(-1.96)	(-0.73)	(-1.04)	(-4.03)	(-1.61)	(-2.96)	(-4.53)
High−Low	0.35	0.22	-0.34	-0.17	-0.48	-1.00	-0.61	-0.87	-1.47
	(1.51)	(0.91)	(-1.51)	(-0.83)	(-2.31)	(-4.27)	(-1.91)	(-3.14)	(-4.01)

Table A4
Portfolios sorted by retail holdings, orthogonalized analyst coverage, and MAX

At the end of each month, retail stocks are partitioned into three groups with tercile breakpoints based on an ascending sort of the orthogonal component of the analyst coverage ($\mathrm{CVRG}^{size\perp}$) and quintiles based on an ascending sort of MAX. This table presents the FF5 alphas for each of the 15 portfolios obtained from the bivariate portfolio sorts. The last row reports the return difference between quintile 5 (High) and quintile 1 (Low) MAX-based portfolios within each attention-based group. Newey-West adjusted t-statistics are given in parentheses.

MAX	Low $\mathrm{CVRG}^{size\perp}$	Medium	High $\mathrm{CVRG}^{size\perp}$
Low	0.00	0.13	0.07
	(0.00)	(1.05)	(0.41)
2	-0.11	0.21	0.10
	(-0.75)	(1.37)	(0.53)
3	0.22	-0.06	-0.46
	(1.37)	(-0.27)	(-2.61)
4	-0.20	-0.29	-0.31
	(-0.91)	(-1.35)	(-1.10)
High	-0.56	-0.78	-1.45
	(-1.64)	(-2.29)	(-5.49)
High−Low	-0.56	-0.91	-1.51
	(-1.51)	(-2.46)	(-4.92)

Table A5
Portfolios sorted by retail holdings, social interactions, and MAX

At the end of each month, stocks in the sample are divided into three groups with tercile breakpoints (30%−40%−30%) based on an ascending sort of retail holdings (RHLD). Stocks are also independently partitioned into three groups with tercile breakpoints based on an ascending sort of a social interaction variable (SOCIAL) and quintiles based on an ascending sort of MAX. The three-way portfolio sorts generate a total of 45 value-weighted portfolios. Each panel of this table presents the FF5 alphas for each of the 45 portfolios. The last row reports the return difference between quintile 5 (High) and quintile 1 (Low) MAX-based portfolios within each social-interaction-based group and each RHLD-based group. The headquarters' social interaction variables are population density (PD) in Panel A, and the Facebook social connectedness index (SCIH) in Panel B. Newey-West adjusted t-statistics are given in parentheses.

Panel A. PD as the proxy for social interactions

	Low RHLD			Medium RHLD			High RHLD		
MAX	Low PD	Medium	High PD	Low PD	Medium	High PD	Low PD	Medium	High PD
Low	0.02	0.03	0.00	0.00	0.11	0.20	0.17	0.14	0.07
	(0.16)	(0.34)	(0.03)	(-0.03)	(1.25)	(2.03)	(1.26)	(1.09)	(0.50)
2	-0.18	-0.04	0.00	-0.09	0.07	-0.26	0.05	0.31	-0.18
	(-1.61)	(-0.35)	(-0.01)	(-0.62)	(0.55)	(-2.27)	(0.32)	(1.51)	(-0.93)
3	-0.09	0.10	0.11	0.05	0.12	-0.15	0.14	0.11	0.29
	(-0.68)	(0.83)	(0.92)	(0.26)	(0.77)	(-1.01)	(0.82)	(0.45)	(1.16)
4	0.10	0.34	-0.18	-0.30	-0.10	-0.31	-0.37	0.12	-0.42
	(0.62)	(2.39)	(-1.19)	(-1.86)	(-0.58)	(-1.89)	(-1.65)	(0.56)	(-1.79)
High	0.07	0.30	-0.25	-0.40	-0.40	-0.41	-0.49	-0.73	-1.29
	(0.38)	(1.79)	(-1.52)	(-1.52)	(-2.13)	(-1.63)	(-1.88)	(-2.13)	(-4.71)
High−Low	0.06	0.26	-0.26	-0.39	-0.51	-0.61	-0.66	-0.87	-1.35
	(0.24)	(1.27)	(-1.16)	(-1.34)	(-2.41)	(-2.06)	(-2.55)	(-2.35)	(-4.73)

Panel B. SCIH as the proxy for social interactions

	Low RHLD			Medium RHLD			High RHLD		
MAX	Low SCIH	Medium	High SCIH	Low SCIH	Medium	High SCIH	Low SCIH	Medium	High SCIH
Low	0.13	0.04	-0.15	0.05	0.19	0.19	0.27	0.22	-0.02
	(1.12)	(0.43)	(-1.40)	(0.35)	(2.22)	(2.06)	(1.93)	(1.67)	(-0.14)
2	-0.20	0.00	-0.08	0.14	-0.09	-0.01	0.18	0.07	-0.24
	(-1.68)	(-0.05)	(-0.79)	(1.13)	(-0.83)	(-0.11)	(1.16)	(0.38)	(-1.27)
3	0.10	0.00	0.07	0.23	-0.09	-0.11	0.03	0.02	0.26
	(0.84)	(-0.02)	(0.69)	(1.42)	(-0.75)	(-0.60)	(0.20)	(0.07)	(1.08)
4	0.02	0.33	-0.33	-0.24	-0.08	-0.40	-0.43	0.01	-0.27
	(0.10)	(2.51)	(-2.34)	(-1.30)	(-0.46)	(-2.50)	(-2.34)	(0.05)	(-1.06)
High	0.26	0.26	-0.28	-0.32	-0.26	-0.34	-0.64	-0.70	-1.43
	(1.44)	(1.58)	(-1.54)	(-1.51)	(-1.68)	(-1.47)	(-3.01)	(-2.22)	(-4.39)
High−Low	0.13	0.22	-0.13	-0.37	-0.45	-0.53	-0.91	-0.91	-1.41
	(0.58)	(1.04)	(-0.56)	(-1.39)	(-2.59)	(-1.98)	(-3.33)	(-2.64)	(-4.09)

Table A6
Fama-MacBeth regressions using size-orthogonalized retail ownership

This table reports the time series averages of the slope coefficients obtained from monthly ordinary Fama-MacBeth regressions of monthly excess returns (in percentages) on lagged MAX and its interaction with a lagged investor attention proxy (ATNT) or social interaction proxy (SOCIAL) and a retail holding dummy. Retail ownership is orthogonalized relative to firm size by regressing RHLD on the natural logarithm of market capitalization monthly. D^H, D^M, and D^L are the retail holding dummy variables defined based on the regression residuals, set equal to one if a stock's residual RHLD is in the top tercile, medium tercile, and bottom tercile RHLD-based groups, respectively, and zero otherwise. The proxies for investor attention are natural logarithm of the number of analysts covering a stock (LNCVRG), absolute value of the standardized quarterly unexpected earnings (|SUE|), and the recency of the lottery event (RECENCY). The proxies for a firm's headquarters' social interactions are population density (PD) and Facebook social connectivity (SCIH). The control variables include market beta (β^{MKT}), the natural log of market capitalization (SIZE), the natural logarithm of book-to-market (BM), momentum (MOM), illiquidity (ILLIQ), co-skewness (COSKEW), standardized unexpected earnings (SUE), and idiosyncratic volatility (IVOL). The slope coefficients of the PD-related interaction terms are multiplied by 100. The last two rows report the average number of monthly observations used in the Fama-MacBeth regressions and the average adjusted R-squared from the Fama-MacBeth regressions, respectively. Newey-West adjusted t-statistics are given in parentheses.

	Attention proxies			Social interaction proxies	
	LNCVRG	|SUE|	RECENCY	PD	SCIH
variable	(1)	(2)	(3)	(4)	(5)
MAX×ATNT×D^H	-0.0185	-0.0472	-0.2389	-0.1205	-0.0677
	(-2.32)	(-4.85)	(-6.95)	(-1.81)	(-4.03)
MAX×ATNT×D^M	-0.0142	-0.0197	-0.1549	-0.0961	-0.0212
	(-2.15)	(-2.26)	(-5.48)	(-1.74)	(-1.36)
MAX×ATNT×D^M	0.0046	-0.0031	-0.0713	-0.1104	-0.0074
	(0.64)	(-0.38)	(-2.38)	(-0.72)	(-0.51)
Controls			MAX, ATNT/SOCIAL, RHLD, β^{MKT},SIZE BM, MOM, ILLIQ, COSKEW, SUE, IVOL		
N	2,112	2,462	2,462	2,380	2,404
Adj. R^2	0.19	0.17	0.17	0.17	0.17

Table A7

Fama-MacBeth regressions using retail stocks: Investor attention and lottery stock returns

This table reports the time series averages of the slope coefficients obtained from regressing monthly excess returns (in percentages) for retail stocks (i.e., stocks in the top tercile retail ownership group) on a lagged MAX and its interaction with a lagged attention proxy (ATNT) or a social interaction proxy. The proxies for investor attention are the natural logarithm of the number of analysts covering a stock (LNCVRG), the absolute value of the standardized quarterly unexpected earnings (|SUE|), and the recency of the lottery event (RECENCY). The social interaction proxies are headquarters' population density (PD) and Facebook social connectedness index (SCIH). The control variables include market beta (β^{MKT}), the natural log of market capitalization (SIZE), the natural logarithm of book-to-market (BM), momentum (MOM), illiquidity (ILLIQ), co-skewness (COSKEW), standardized unexpected earnings (SUE), idiosyncratic volatility (IVOL), and quarterly retail ownership (RHLD). The slope coefficients of the PD-related interaction terms are multiplied by 100. Newey-West adjusted t-statistics are given in parentheses.

Variable	ATNT=LNCVRG	ATNT=|SUE|	ATNT=RECENCY	SOCIAL=PD	SOCIAL=SCIH
MAX×ATNT/SOCIAL	-0.0239	-0.0435	-0.1447	-0.1066	-0.0576
	(-2.14)	(-3.67)	(-3.06)	(-0.96)	(-2.51)
MAX	-0.0340	-0.0273	-0.0176	-0.0456	-0.0358
	(-1.74)	(-1.83)	(-1.13)	(-2.69)	(-2.30)
ATNT/SOCIAL	0.1425	0.2091	-0.3111	0.0015	0.2114
	(1.93)	(2.99)	(-1.24)	(0.31)	(1.75)
β^{MKT}	0.0361	0.0586	0.0285	0.0074	0.0174
	(0.37)	(0.64)	(0.29)	(0.07)	(0.18)
SIZE	-0.2363	-0.1653	-0.1602	-0.1573	-0.1569
	(-4.63)	(-4.83)	(-4.27)	(-3.91)	(-4.08)
BM	0.1909	0.2091	0.2222	0.2435	0.2281
	(2.49)	(3.47)	(3.65)	(3.98)	(3.73)
MOM	0.0054	0.0046	0.0044	0.0050	0.0044
	(3.96)	(4.59)	(4.23)	(4.86)	(4.06)
ILLIQ	-0.0532	-0.0040	0.0080	0.0023	0.0022
	(-2.83)	(-0.59)	(1.04)	(0.28)	(0.29)
COSKEW	-0.2028	-0.2591	-0.2999	-0.2946	-0.3121
	(-1.24)	(-1.79)	(-1.77)	(-1.70)	(-1.94)
SUE	0.4511	0.5017	0.5447	0.5781	0.5493
	(13.73)	(17.92)	(14.52)	(16.47)	(15.20)
IVOL	-0.2438	-0.1758	-0.2227	-0.1952	-0.1739
	(-3.73)	(-3.07)	(-3.47)	(-2.92)	(-2.78)
RHLD	-0.0033	-0.0091	-0.0222	-0.0012	-0.0037
	(-0.52)	(-1.83)	(-1.23)	(-0.10)	(-0.31)
N	503	735	529	520	532
Adj. R^2	0.15	0.13	0.13	0.13	0.13

Table A8
Summary statistics for retail order imbalances

This table compares the summary statistics for daily retail order imbalances between our sample (BHPT) and the sample of Boehmer et al. (2021) (BJZZ). The retail order imbalances are measured based on share volume (OIBVOL) and the number of trades (OIBTRD). Our sample covers the period January 2010 through December 2017. The BJZZ sample covers the period of January 2010 through December 2015.

	Mean	Std Dev	Median	25^{th} Pctl	75^{th} Pctl
OIBVOL (BHPT)	-0.026	0.398	-0.024	-0.292	0.234
OIBVOL (BJZZ)	-0.038	0.464	-0.027	-0.301	0.217
OIBTRD (BHPT)	-0.018	0.319	0.000	-0.224	0.197
OIBTRD (BJZZ)	-0.032	0.437	-0.010	-0.276	0.205