Reddit v. Wall Street: Why Redditors beat Wall Street at its own game

Nishita Bhakar¹, Logan Pearce²
¹ Redmond High School, Redmond, Washington
² Princeton University, Princeton, New Jersey

SUMMARY
In January of 2021, users of the internet forum Reddit, known as Redditors, caused a short squeeze of GameStop that led to major losses for short sellers. In our research, we sought to find potential incentives behind the sudden increase in GameStop stock price during the start of 2021. We hypothesized that those who were involved in this event spent large sums of money to purchase stock because they wanted to fight back against Wall Street, who they perceived as the “big guys” who had been controlling the stock market for a long time. Using a combination of methods, we showed that their true motivations may not have been strictly financial. Through qualitative analysis, we found that after the peak in GME stock prices, Redditors generally expressed positive emotions towards the short squeeze. Quantitative analysis revealed that although Redditors initially used neutral language towards the short squeeze, they increasingly used positive language. Thus, our research suggests that Redditors may have initially been financially motivated to short-squeeze GameStop stock, but as time progressed, emotional motivations may have become more important. Our work shows that social phenomena can be dynamic and evolve. It also highlights the value of using a mixed methods approach to study complex processes in the real world. It is important to note that we were only able to learn the motivations of Redditors who posted on the r/WallStreetBets community, and though they had a significant influence on the short squeeze, they were realistically not the only shareholders, meaning our results are not applicable for all the shareholders.

INTRODUCTION
In January 2021, a group of investors on the internet forum Reddit, known as Redditors, rallied to buy GameStop stock causing the stock price to increase rapidly and leaving some Wall Street hedge funds with billions in losses (1). Wall Street has been the heart of financial activity since 1792 (2). It is composed of a variety of financial markets, including those for stocks, bonds, and foreign exchanges (2). While originally Wall Street was created primarily to raise capital to support companies, securities trading has become a lucrative profession by itself (2). One lucrative and risky process is “short selling” a stock. In essence, professionals on Wall Street who “short” make a bet that a certain company will lose money over some period of time. If the company does indeed do poorly, then the professional investors win the bet and make money; if the company actually does better, then the investors lose the bet and face potentially enormous losses (3).

In late 2020, many people bet that GameStop stock would do poorly since GameStop was struggling severely at the time. The company announced in September 2020 that they were closing about 100 more stores than the 320 that they planned to close in March 2020 (4). Redditors on the “subreddit” community r/WallStreetBets (WSB) saw the opportunity for a short squeeze (5). A short squeeze is the process of buying lots of shares of a company’s stock, thereby increasing the price and costing the short sellers substantial sums of money (3). In this work, we explored the underlying incentives of Redditors beating Wall Street. There are many potential explanations. Historically, white males have dominated Wall Street, a trend that continues to this day. Wall Street is still characterized by racial and gender inequality (6). Many people think of Wall Street as the “big guy” taking advantage of the “little guy,” so Redditors may have seen this as their opportunity to switch that narrative. Alternatively, perhaps Redditors simply thought of it as a game, something to occupy their time in the monotony of a global pandemic and remote work. Our research question was to identify the potential underlying incentives as to why the WSB Redditors short-squeezed GameStop (GME) stock. Whatever the reasons, we knew that the key to answering our research question was diving deep into WSB.

To delve into WSB, we combined qualitative and quantitative methods. We used grounded theory analysis to understand the nuances of 45 popular posts made after the peak of GameStop stock prices. Grounded theory analysis is an iterative technique used to understand the relationships between qualitative data, such as interviews (7). We used this technique to categorize the posts and identify patterns. Through this technique, we found that Redditors generally supported the short squeeze. Those who did not support the short squeeze did so out of concern for the Redditors involved. We also used sentiment analysis, a quantitative process that identifies whether the author’s opinion towards a topic is positive, negative, or neutral, to examine a larger number of posts. The quantitative analysis paralleled our qualitative analysis; while initially the opinions were neutral,
they became more positive as time went on. We concluded that while initially Redditors short-squeezed the stock for financial motives, their emotional involvement became the driving factor over time. We also note that it is impossible to make causal claims from correlational data. However, by combining multiple methods, we can better understand the phenomena that we are studying and make stronger claims. Our work also underscores that social movements can involve dynamic processes that evolve over time. Studying data that exists naturally without researcher interference, such as Reddit, allows us to understand these movements in a powerful way.

RESULTS
Timeline of GME-Related Activity on WSB
We used the Pushshift API to scrape data from WSB from April 26, 2012 to April 27, 2021. “Scraping” means that we used computer programming to collect posts from WSB over a certain time frame. The details on how the Pushshift API works are explained in the Methods section. Using all this data, we plotted the total number of posts starting from May 2018 on a timeline, which allowed us to see temporal trends of activity on WSB. We determined that activity on WSB started to pick up around the start of 2020 after over 5 years of little to no activity (Figure 1). We also observed that at the end of 2020 to early 2021 there was a huge surge in activity, with the number of posts exceeding eight thousand. This spike in popularity coincides with the start and end of the GME phenomenon.

We hypothesized the observed increase in posts on WSB was due to the GME phenomenon. To test this hypothesis, we looked at the total number of posts from all our scraped data, starting from May 2018, that included keywords “GME” or “GameStop” in their title. Similar to the total number of posts on WSB, there was increased activity from the start of January 2021, which returned to baseline levels by the end of February 2021 (Figure 1). Since we only used data from the “GME” and “GameStop” subset, it allowed us to support our hypothesis that the sudden boom in activity on WSB was related to the GME phenomenon. We also observed that prior to October 2020 the number of posts mentioning “GME” or “GameStop” was less than 50, but something happened in October 2020 that spurred an increase in the number of posts, exceeding 1750 by the end of January 2021 (Figure 1).

Timeline of Popular Words and Phrases on WSB
To further explore the data, we used word clouds to give us access to new types of information, such as finding when the phrase “GME” and/or “GameStop” became popular. We created word clouds using all the scraped data for each month from July 2020 to March 2021 using words from both the title and the post itself (Figure 2).

The most popular phrases in the word clouds for July, August, and September of 2020 pertained to common stock market slang and Tesla, Inc., an electric car company. However, in October 2020, the phrase “GME” suddenly spiked in popularity and was the second-largest word in the word cloud for October 2020 (Figure 2). This information gave us clear direction in pinpointing when the GME phenomenon started. From October 2020 to March 2021, GME stayed on the word clouds’ five most-used words list, which further solidified our period for the GME event.

Figure 1: Activity on the WSB subreddit. The line graph shows that there was a notable peak in the number of posts on WSB between January and February of 2021.
Additionally, we used a one-tailed heteroscedastic t-test to compare the number of posts containing “GameStop” or “GME” per week in weeks stretching from June 11, 2018 to December 28, 2020 versus the weeks from January 4, 2021 to February 1, 2021, and found that the difference was not statistically significant (t-test, \( p = 0.0673 \)). However, the actual difference is likely more significant due to gaps in the weeks. We were only analyzing posts that contained “GameStop” or “GME,” and some weeks did not have any posts that matched these criteria. Thus, there are a few weeks from June 6, 2018 to October 5, 2020 that don’t have any posts. It is not until October 10, 2020 and onwards that there was at least one post per week (Table 1).

## Identifying and Analyzing the Original Post

Next, we identified the original post that may have ignited the Redditors to game Wall Street. As previously discussed, we noticed the phrase “GME” started to show up in posts on WSB around the start of October, so we began our search for the original post on the week of October 5, 2020 since it was the first week that showed an increase in the number of posts containing “GME” (Table 1, highlighted). These 16 posts were already discussing short squeezing the GME stocks, suggesting none of them were the original post. We identified a post on September 14, 2020 written by the Reddit user FatAspirations that was titled “Cheap GME calls for a potential lottery win” with 373 upvotes and 489 comments at the time of our analysis. This post, reproduced below, discussed the opportunity to profit from the many people who were short selling GME stock (8).

> “Gonna keep this short and simple.
> There’s some funky things going on with Gamestop (GME), a.k.a. the blockbuster of video games.
> 1. Short fee >30% and rising
> 2. Short availability <200K shares and falling
> 3. The selloff the last week in GME was largely correlated with overall market selloff. IF there’s a rebound this week, possible to see rebound in GME as well.
> If the stock pops any more than it did today, there may be a lot of people rushing to cover their very expensive shorts. Positions:
> 9/18 GME 6c, 7c, 7.5c, 8c”
We thus concluded this to be the original post that likely incited other Redditors to become involved in investing in GME stock. The Reddit user also included 20 updates to the original post (see Appendix). These updates mainly discussed changes in the short fee, details about where and how much the other Redditors should invest, as well as encouragement to continue investing. We also note that this post and others in our analyses may have since been upvoted/downvoted and some comments have been added, removed, or changed.

Analyzing Other Posts

FatAspirations’ post gave us a starting point from where we could formulate the motivations behind the Redditors who sought to beat Wall Street at its own game. Since it would be infeasible to read each individual post in detail, we chose to focus our attention on posts by certain Redditors. To pick the Redditors, we created a table depicting the top 10 users ordered by the number of posts they had made on WSB (Table 2). One thing to note is the “[deleted]” post count does not refer to a specific Reddit user; it is the sum of all the posts of users who had deleted their accounts.

For feasibility, we focused on two Redditors to study. We wanted to choose these two users strategically so that their posts covered the time before, during, and after the GME event. We created three word clouds to show the most active users during September of 2020, the month of the original GME post by FatAspirations (Figure 3A), the most active users during January of 2021, the month containing the first half of the GME event’s peak (Figure 3B), and the most active users during February of 2021, the month containing the second half of the GME event’s peak (Figure 3C).

The two users we chose study were FatAspirations and HomeDepotHank69. These two Redditors’ posts would allow us to feasibly analyze diverse and prolonged data on WSB. We chose FatAspirations, whose posts stretched from September 14, 2020, to February 7, 2021, because their username was prominent on the first two word clouds, allowing us to get data before and during the GME phenomenon. We also chose HomeDepotHank69, whose posts stretched from February 26, 2021, to April 27, 2021, since they were the user with the largest number of posts on WSB, which allowed us to study posts after the GME event (Table 2).

These users had a combined total of 21 posts, 11 of which were relevant to our study. We examined these posts for two main features: changes in opinions and overall trends. First, we studied FatAspirations’ posts (8). In FatAspirations’ initial posts, titled “Cheap GME calls for a potential lottery win” and “GME - EndGame: DTC Infinity.,” we noticed that they mainly focused on the raw data regarding GME stocks, such as the increase and decrease in short fees and the prices of GME shares, with a profit-centered point of view (9, 10). The other WSB Redditors were also focused on the financial aspect of GME stock.
the event; the comment sections of FatAspirations’ initial posts were filled with data discussion and the opportunity to earn a large profit. However, a striking result of our analysis was that the opinions and motivations of the Redditors appeared to change extremely rapidly.

Unlike FatAspirations’ previous posts, in their third post, titled “GME - EndGame Part 2: Cohen, Market Cap, Potential Investors”, rather than describing the data surrounding GME, they discussed that this was a once-in-a-lifetime opportunity that Redditors should not miss out on: “If you’re not in GME yet, ask yourself how you would truly feel if what everyone here is saying panned out to be true, and you weren’t participating” (11). The titles of the subsequent posts by FatAspirations encapsulate the emotions expressed by the Redditors in their posts – “GME EndGame part 3: A new opponent enters the ring”; “GME - EndGame Part 4: The Saga Continues”; “GME - EndGame Part 5: They couldn’t win, so they cheated. Why you should leave Robinhood.”; and “GME - EndGame Part 6: The Big Reset, or The Greatest Financial Crime of the Century - and how to play GME going forward” (12-15). In these titles, FatAspirations portrays the GME event less as a short squeeze and more as an act of defiance, leading us to believe that the most likely potential incentive behind the Redditors’ later actions was perceiving an unjust hierarchy in the stock market.

To get another perspective on the GME event, we then examined HomeDepotHank69’s posts, which focused on the analytical side of GME and eschewed emotion (16). HomeDepotHank69, along with a few other Redditors, were focused on building a strategy to maximize their profits; however, it seems that they were not doing this for selfish reasons. In fact, many Redditors in the comment section of these posts discussed how they planned to donate their profits to charities, revealing the primary incentive may have not been individual financial gain. The Redditors of WSB also came together to donate over $350,000 to the Dian Fossey Gorilla Fund that protects and studies gorillas; this further underscores this idea of selflessness (17).

### Grounded Theory and Sentiment Analysis

After studying all the scraped data from WSB, we used grounded theory to analyze a subset of the data; specifically, 45 posts from January 2021 with a popularity score greater than 33. We looked for overall trends between the posts, such as overarching opinions, similar authors, specific concerns regarding the short squeeze, etc. We found that 84% of posts supported the short squeezers. Our definition for a post supporting the short squeeze was if the author took part in the short squeeze or if they spoke positively about it. The remaining non-supporting posts primarily expressed their concern towards the Redditors who had invested large sums of money and advised them to be wary of how much they invested. We also noticed that most posts could be put into one of four categories; we developed these categories using grounded theory analysis. The four categories were summarizing recent events in GME stocks, planning where to invest in future, analyzing data and stock trends, and warning investors of upcoming trends. Also, the posts which were the
most liked, were those that either analyzed the stock trends or those that advised where the Redditors should invest next.

Though grounded theory allowed us to identify the nuances and relationships between individual posts, we were unable to go through every single post surrounding the GME phenomenon with such an intensive process. Thus, we chose to apply sentiment analysis on the full dataset to process greater amounts of information. From the results, we learned that the Redditors overall initially had a neutral viewpoint towards investing in GME stocks with sentiment scores averaging at 0, but as time progressed, the sentiment scores increased rapidly from 0 towards 1, a strong positive opinion. This trend continued after the GME phenomenon’s peak; since February 2021 to April 2021, the sentiment score continued to stay close to 1, which means the text showed an overall positive opinion. This information aligns with our previous findings from looking into specific posts.

DISCUSSION

Using a variety of both quantitative and qualitative methods, we identified potential underlying motivations of the Redditors who rattled Wall Street with their GameStop stock purchasing decisions. Through our initial data collection and word clouds, we were able to identify the exact date that the phenomenon likely began. Many news outlets speculated that the event started as an emotional one led by “a bunch of hyper-online mischief-makers in Reddit’s r/WallStreetBets forum—a clan of self-described degenerates” (18). However, by analyzing this original post, we were surprised to discover that the initial motivations behind gaming Wall Street may have been simply financial, because FatAspirations appeared to be initially focused solely on the financial benefit of investing in GME. However, as time progressed, the emotional investment in manipulating Wall Street may have increased. In FatAspirations’ third post, they highlighted the importance of emotional conviction in the endeavor. In this post, FatAspirations asks Redditors how they would “feel” if “everyone else” besides them participated in the event and the event worked out. This post may have incentivized emotionally charged Redditors to buy their own GME shares. Additionally, the shift from neutral to increasingly positive scores from our sentiment analysis suggests that emotions became increasingly relevant.

While the primary motivations behind the short squeeze may have been financial, it is also interesting to note that none of the posts we found objected to the short squeeze from a moral perspective. These objections match our initial hypothesis that the “little guys” were tired of the “big guys” on Wall Street. In other words, the Redditors did not object because they felt that the Redditors were unfairly taking advantage of Wall Street; they objected because they were concerned for those who were doing the short squeeze. However, it is possible that there were Redditors on the WSB subreddit who did not post their opinions or did so outside of the subreddit. Taken together, the entire event demonstrates the impact one individual may have had and the potentially shifting, complex nature of motivations.

We used a variety of both quantitative and qualitative methods to better answer our research question of why the Redditors on WSB manipulated GameStop stock. Both analytical methods have their strengths and limitations. Through sentiment analysis, we were able to analyze every single post in WSB. However, sentiment analysis only yields a number and hides the words that lead to that sentiment score. Through grounded theory analysis and in-depth examinations of individual posts, we were able to identify rich nuances that were not expressed simply through numbers. Thus, the breadth of methods we used complemented each other and allowed us to approach the same question from many different angles.

There were limitations that our study faced due to feasibility constraints and the amount of time that we invested into deeply analyzing each post. It was not possible for us to study each post closely with over 40,000 posts, each with an average of over 3,000 words. To mitigate this issue, sentiment analysis allowed us to quickly study each post in WSB. An additional limitation is that we only studied WSB, not the GME subreddit. However, the WSB subreddit was much more relevant to our research question since news outlets had already reported that the groundwork began there (19).

This paper provides an important example of how to study sociological phenomena in a real-world context. Future studies should similarly seek to combine quantitative and qualitative methods. For example, researchers can investigate why individuals work together to make social change by using sentiment and grounded theory analysis on Reddit and other forums.

MATERIALS AND METHODS

We used the Pushshift API to scrape the posts on WSB (20). Pushshift allows one to scrape Reddit posts by specifying the subreddit, start and end date, and other metadata for posts. However, one downside to using Pushshift API is that it cannot scrape previously deleted posts, but it can scrape posts from deleted users. We chose to collect the author, title, date, score, and number of comments for each post from WSB from April 26, 2012 (the start date of WSB) to April 27, 2021 (the date we began our research and scraped WSB). The score of a post measures its relative popularity; it is computed by subtracting the total number of “downvotes” from the total number of “upvotes” and then running this value through the proprietary Pushshift API algorithm. We chose to focus on these five pieces of metadata per post since they were the most relevant to answering our research question.

Although the peak of the GameStop stock price occurred between late 2020 and early 2021, we collected data starting from 2012 to determine what potentially caused the first spark in the GME phenomenon. April 27, 2021 was a good time to scrape the data as the GME stock prices peaked on
January 29, 2021; thus, we were able to analyze the subreddit both during and after the crux of the GME phenomenon (Figure 4) (21). While collecting the raw data, we also created a subcategory that held all the posts from WSB containing the term “GME” or “GameStop” in their title. This subcategory allowed us to pinpoint the exact time when GME started to gain popularity in the WSB community as well as conduct extensive analysis on particular sections of our data.

To visually explore all the data that we had collected through scraping, we created several types of graphs and tables. We constructed line graphs plotting the number of posts on a timeline, tables showing the activity of various users, and word clouds depicting the most-used words in posts during a specified time using the Python tool plotly. Our word clouds identified and displayed the five hundred most-used words from the full dataset. The sizes of the words represent how often they were used. In our first test runs, we noticed that filler words (e.g., like, so, the) took up most of the output. To avoid this issue, we used a tool called Stopword that automatically removed specified words from the data. The words we chose to remove were the, of, so, and, a, to, in, it, was, are, r, you, u, my, I, and like. These graphs cumulatively allowed us to pinpoint the exact period the GME stock started to gain popularity and in turn find the original post that appeared to rally Redditors to take part in a short squeeze.

We also employed grounded theory analysis. We looked at individual Reddit posts and grouped posts with similar themes together. After initially categorizing the posts, we continued to re-group the posts until we identified four categories: posts that summarized recent events in GME stocks, planned where to invest in future, analyzed data and stock trends, and warned investors of upcoming trends. This process allowed us to identify patterns within the data and see how posts relate to each other. Our initial data from scraping yielded over 420,000 posts, which was clearly infeasible to go through by hand. Instead, we focused on all the posts with a popularity score greater than 33 in January 2021, the first half of the GME phenomenon’s peak. We focused on the first half because we believed that the Redditors’ opinions would be strongest when the GME stock prices were increasing rapidly (21). We chose a cutoff for scores greater than 33 to create a feasible number of posts to go through by hand and to ensure that all the posts were relevant to the GME phenomenon. A lower cutoff would have led to too many posts to analyze in depth. Additionally, those posts may have been less popular because they were irrelevant to the GME phenomenon. A higher cutoff would have decreased the number of posts we analyzed, potentially causing us to draw inaccurate conclusions based on a few extremely popular posts. In total, we used grounded theory on 45 posts.

In terms of quantitative methods, we used Google Sheets to conduct a one-tailed heteroscedastic t-test that compared the number of posts containing “GameStop” or “GME” per week in weeks stretching from June 11, 2018 to December 28, 2020 versus the weeks from January 4, 2021 to February 1, 2021. Our significance level was the standard level of $p < .05$. We also used Python 3.10.0 to implement sentiment analysis on all posts to see the popular opinions among Redditors during different periods of time regarding the GME stock prices. We inserted each post into the Natural Language Toolkit, which uses natural language processing to yield a compound sentiment score for each post (22). The compound score represents the overall feeling of the text; how positive, negative, or neutral it is. It ranges between -1 and 1; if the compound score is closer to -1 than 1, the author has an overall negative opinion towards the topic and vice versa.
REFERENCES


Copyright: © 2022 Bhakar and Pearce. All JEI articles are distributed under the attribution non-commercial, no derivative license (http://creativecommons.org/licenses/by-nc-nd/3.0/). This means that anyone is free to share, copy and distribute an unaltered article for non-commercial purposes provided the original author and source is credited.
Appendix

The original post from r/FatAspirations (including updates) is reproduced here:

Gonna keep this short and simple.
There's some funky things going on with Gamestop (GME), a.k.a. the blockbuster of video games.

1. Short fee > 30% and rising
2. Short availability <200K shares and falling
3. The selloff the last week in GME was largely correlated with overall market selloff. IF there's a rebound this week, possible to see rebound in GME as well.

If the stock pops any more than it did today, there may be a lot of people rushing to cover their very expensive shorts.

Positions:
9/18 GME 6c, 7c, 7.5c, 8c

Update: Adding to positions on dips. https://imgur.com/a/oKTV4ZV
Update 2: Party is getting started. Up 50% in 2 hours. Jump on before it's too late!
Update 3: Ended day up 72%. Still holding, still going up AH.
Update 4: The short fee has increased to >38%, with 0 shares available to short. GME is over 100% shorted now!
Update 5: The short fee is now over 50%.
Update 6: Been adding to my positions on dips, but there haven't been many today. Haven't sold any thing yet. Up to 118 contracts. Current positions:
Update 7: Still holding/adding. Short fee now at 64%. I'm speculating, but imagine paying 64% annual interest to short the stock while it goes up 30% in 2 days.
Update 8: Buying every dip (major red candle at 2:30 on 9/17)
Update 9: Hope you are buying these dips!
Update 10: here's what I'm sitting on RH. https://imgur.com/a/WCjYRgc Most positions are 9/18 calls. Also have about 4K shares and another 70 contracts in IBKR. Largest single position in terms of qty is 9/18 10c - about 80 contracts in that strike. Haven't sold any calls (ITM or OTM) yet.
Update 11: Day of reckoning for my 9/18 holdings. Still holding as of 7:00 a.m.
Update 12: Still holding. Added more 9cs when they went to .2 on the dip. Anyone who wants FDs - 9/18's 9cs are trading at intrinsic value.
Update 13: I hope you’re still holding!!
Update 14: https://fintel.io/ss/us/gme <- More than 10M new shorts in the last 2 trading days, at >50% interest rates, while GME only goes up. I hope you’re holding. ITMs or just OTM for 9/25, 10/16, etc. are good on any dips. or just get shares.
Update 15: Like I said in update 8, buy the 2:30 dip.
Update 16: F[***] RH. They auto-sold my ITM calls. Bought back in on IBKR.
Update 17: Final update for the week. Up $24K on calls/shares, but a lot of OTM calls expired worthless. Hope you did well. I’m going into the weekend holding $65K worth of shares and about $8K worth of 9/25 calls - at strike prices from 6, 7, 7.5, 8, 9, 10, 11. Not going further up than that for now. Lost a good chunk of profits from not knowing RH’s auto-sell process (first month using RH).
Update 18. Still holding / rolling. Buy major dips in the morning. 9/25 10cs were 0.49 the morning of Wed.
Update 19 Got out of most of my positions here. Hope you timed things better than me. At my peak I was up $30k, getting out with a net of $15k.
Update 9/29 currently holding 4500 shares (added 1500 over the last 24 hours) and a few 9/25 calls. Gradually increasing shares on dips