SOCIAL MEDIA CENSORSHIP:
THE IMPACT FILTERING INFORMATION HAS
ON THE WELL-BEING OF USERS AND ONLINE BUSINESSES

AN HONORS THESIS
SUBMITTED ON THE 23\textsuperscript{RD} DAY OF APRIL 2021
TO THE A.B. FREEMAN SCHOOL OF BUSINESS
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
OF THE HONORS PROGRAM
OF NEWCOMB TULANE COLLEGE
TULANE UNIVERSITY
FOR THE DEGREE OF
BACHELOR OF SCIENCE IN MANAGEMENT
WITH HONORS IN BUSINESS

BY

\textit{Katherine Chapin}

KATHERINE ELIZABETH CHAPIN

APPROVED:

\underline{Michael Wilson}

MICHAEL S. WILSON, PhD
Director of Thesis

Mita Sujan

MITA SUJAN, PhD
Second Reader
ABSTRACT

Katherine Chapin. Social Media Censorship: The Impact Filtering Information Has on the Well-Being of Users and Online Businesses

Dr. Michael S. Wilson

This thesis explores the role of social media companies in filtering online users’ content, especially in the case of health-related misinformation outbreaks. Specifically, the thesis discusses the history of censorship policies, the spread of health-related misinformation, the impacts of misinformation on individuals and companies, the most effective means to correct misinformation, and how the spread and correction of misinformation changed during the 2020 to 2021 COVID-19 pandemic. The business case-structured thesis includes three sections: a case study, an instructor’s manual, and a literature review. The case study explores Facebook’s history with filtering information on its online platform, its reactions to misinformation outbreaks, and the company’s response to misinformation during the COVID-19 pandemic. The literature review examines scholarly articles about health-related misinformation, compares and contrasts scholars’ viewpoints about the topic, and concludes with suggested recommendations specific to the COVID-19 pandemic. The instructor’s manual offers a teaching plan for the case and provides specific learning objectives and discussion questions for instructors to use in various management, marketing, and finance courses.

Leading scholars agree that authorities can correct misinformation on social media, and proactive filtering measures can reduce misinformation outbreaks. However, the scholars offer differing options for correcting misinformation on social media and they vary regarding the degree of responsibility they believe social media platforms have to address and correct misinformation.
ACKNOWLEDGEMENTS

I would like to thank all of the people who assisted me in planning, writing, editing, and executing this thesis. My first message of gratitude is to Professor Michael Wilson for all of his support, suggestions, and mentorship throughout my entire thesis process. I would not have been able to complete this thesis in an efficient and comprehensive manner without his continued guidance. Additionally, I would like to thank Professor Mita Sujan for her advice and suggestions for my thesis, especially pertaining to the marketing concepts and organizational structure. Her thoughtful insights were essential in crafting the final analyses included in my thesis. I would also like to extend a sincere thank you to Cynthia Fransen for all of her assistance in editing and organizing my final thesis product. Her assistance encouraged me to think carefully and inquisitively about how I expressed my ideas throughout the thesis.

Next, I would like to express my gratitude towards the faculty and staff at Tulane University for the amazing opportunities they have offered me and for supporting me in writing my Seniors Honors Thesis and studying a breadth of topics in the Undergraduate A.B. Freeman School of Business. Finally, I would like to thank my parents for their continued support through my college career and especially during my senior year at Tulane University. I could not have written this thesis without the support of these individuals, as well as my friends and colleagues.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>ii</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>iii</td>
</tr>
<tr>
<td>TABLE OF CONTENTS</td>
<td>iv</td>
</tr>
<tr>
<td>LIST OF CASE STUDY EXHIBITS</td>
<td>vii</td>
</tr>
<tr>
<td>LIST OF INSTRUCTOR’S MANUAL EXHIBITS</td>
<td>viii</td>
</tr>
<tr>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>Scope of Work</td>
<td>1</td>
</tr>
<tr>
<td>Value of Research</td>
<td>2</td>
</tr>
<tr>
<td>Case Study of Facebook</td>
<td>2</td>
</tr>
<tr>
<td>Literature Review of Scholarly Articles</td>
<td>4</td>
</tr>
<tr>
<td>Instructor’s Manual with Teaching Notes</td>
<td>5</td>
</tr>
<tr>
<td>Conclusion of Thesis</td>
<td>6</td>
</tr>
<tr>
<td>References</td>
<td>7</td>
</tr>
<tr>
<td>CASE STUDY</td>
<td>8</td>
</tr>
<tr>
<td>Facebook’s History</td>
<td>8</td>
</tr>
<tr>
<td>The Face Behind Facebook</td>
<td>8</td>
</tr>
<tr>
<td>The Early Days</td>
<td>9</td>
</tr>
<tr>
<td>The Company’s Unprecedented Growth</td>
<td>10</td>
</tr>
<tr>
<td>Facebook’s Growing Community and Political Influence</td>
<td>11</td>
</tr>
<tr>
<td>2016 Presidential Election</td>
<td>12</td>
</tr>
<tr>
<td>Company’s Response to Misinformation</td>
<td>13</td>
</tr>
<tr>
<td>Disrupting Economic Incentives</td>
<td>14</td>
</tr>
<tr>
<td>Helping People Make Informed Decisions</td>
<td>15</td>
</tr>
<tr>
<td>Building New Products</td>
<td>16</td>
</tr>
</tbody>
</table>
Facebook’s Errors in Post Filtering ................................................................. 17
Company’s Focus on Artificial Intelligence Initiative with New CTO ....... 18
Facebook’s Fight Against False News During the COVID-19 Pandemic ... 20
   Efforts to Resolve COVID-19 Misinformation ........................................ 21
Facebook’s Dilemma Establishing Uniform Information Filtering Criteria . 24
Appendix ........................................................................................................... 25
References ....................................................................................................... 29

LITERATURE REVIEW ...................................................................................... 33
Introduction .................................................................................................... 33
History of Free Speech in the United States .................................................. 35
Evaluation of Internet-Content Censorship Policies ..................................... 37
Definition of Misinformation ......................................................................... 39
Review of Third-Party Regulators of Misinformation .................................... 40
Spread of Health-Related Misinformation .................................................... 41
Impacts of Misinformation on Individuals and Companies ......................... 43
   Individual Effects ........................................................................................ 43
   Company Effects ......................................................................................... 43
Ways to Correct Misinformation ..................................................................... 44
   Tracking Demographics and Educating Communities ............................... 45
   Overcoming Psychological Biases .............................................................. 46
   Evaluating Correction Types ..................................................................... 48
   Considering Topic Familiarity and Information Credibility ......................... 52
      Individual Familiarity ............................................................................ 52
      Source Credibility .................................................................................. 54
Correction of COVID-19 Vaccination Misinformation ................................... 56
Summary ......................................................................................................... 58
<table>
<thead>
<tr>
<th>Exhibit</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exhibit 1</td>
<td>Russian-Purchased False Advertisements in 2016</td>
<td>24</td>
</tr>
<tr>
<td>Exhibit 2</td>
<td>2016 Facebook Adjusted Close Stock Price</td>
<td>25</td>
</tr>
<tr>
<td>Exhibit 3</td>
<td>2017 Facebook Adjusted Close Stock Price</td>
<td>25</td>
</tr>
<tr>
<td>Exhibit 4</td>
<td>2020 Facebook Adjusted Close Stock Price</td>
<td>26</td>
</tr>
<tr>
<td>Exhibit 5</td>
<td>Facebook Monthly Active Users Worldwide</td>
<td>26</td>
</tr>
<tr>
<td>Exhibit 6</td>
<td>Facebook Revenue from 2014-2019</td>
<td>27</td>
</tr>
<tr>
<td>Exhibit 7</td>
<td>Confounding Variables in Facebook’s 2020 Stock Price</td>
<td>27</td>
</tr>
</tbody>
</table>
# LIST OF INSTRUCTOR’S MANUAL EXHIBITS

<table>
<thead>
<tr>
<th>Exhibit</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Student Grading Rubric</td>
<td>81</td>
</tr>
<tr>
<td>2</td>
<td>Student Answer Scoring Rubric</td>
<td>82</td>
</tr>
</tbody>
</table>
INTRODUCTION

This thesis analyzes the corporate social responsibility, or lack thereof, of social media platforms to filter misinformation, and offers perspectives about how these efforts impact the brand images and brand equities of these companies.

The spread of misinformation on social media is concerning because many individuals rely on social media platforms as their primary news sources. As consumers increasingly look for convenience in every aspect of their lives, they want to acquire knowledge efficiently and easily. In response to these consumer trends, social media websites have grown tremendously and, increasingly, they are functioning as multi-dimensional platforms. However, with this expanded accessibility also comes a high risk that the information and news that appear on these platforms may be inaccurate. To address this issue and retain credibility, these websites must determine to what extent they have the responsibility to protect their users while maintaining their brand identities, which could potentially have a negative impact on their financial positions. This thesis informs business students about the subject of misinformation on social media platforms and will answer the question: *What is the impact of censoring misinformation and disinformation on social media platforms, both for consumers and for the success of these businesses?*

Scope of Work

This thesis contains three sections: a case study, a literature review, and an accompanying instructor’s manual. The case study focuses on a particular example of a social media platform to explore the larger research question of the thesis. Specifically, the thesis examines Facebook and evaluates the company’s position in 2020 and 2021 during the COVID-19 pandemic and the various challenges it faced as a multi-dimensional social media platform. Facebook had to balance its obligation to filter users’ posts to remain a reliable source of information while at the same time adhere to the business model the company had built around increasing users’ online activities.
The literature review uses insights from scholarly, peer-reviewed articles to discuss the impact that online misinformation, and health-related misinformation, specifically, had both on community health and on the success of social media platforms, in general. The instructor’s manual functions as both an outline and a guide for instructors to use as they lead students’ discussions regarding the role that censoring plays in creating successful, reliable social media platforms.

**Value of Research**

The importance of censoring misinformation is a valuable research topic in societies that are hyper-focused on finding convenience in daily tasks that require trustworthy information sources. As consumers hope to simplify their lives through the use of technology, they look to information platforms that can offer them a range of services, from posting personal photographs to reading the latest news updates. In this information-centric environment, social media companies face the problem of increasing their user activity by being vehicles for free speech while also ensuring that the content on their platforms does not harm consumers by including false or misleading information. Consumers must be able to rely on the accuracy of the information they are receiving and must have confidence in the businesses that are providing this information. Therefore, these platforms must be proactive in filtering misinformation. However, they must do so accurately to ensure that they are not impinging on users’ rights to freedom of speech by deleting opinion pieces or omitting factually correct information.

**Case Study of Facebook**

The case study focuses on a specific company, Facebook, as it deals with the challenge of how to censor information to ensure the reliability of content on its platform during a viral pandemic. The case study evaluates the decision-making process of Facebook’s Chief Technology Officer, Mike Schroepfer, as he attempted to present Facebook’s platform as an interactive and reliable source for personal and global news. The case study examines the strategic efforts of Schroepfer that altered the way in which Facebook evaluated the user content on its site.
This case study begins in February of 2004 when founder Mark Zuckerberg launched Facebook, and it continues through the end of 2020. Schroepfer entered the national spotlight in 2019 when he publicly responded to a gunman who had used Facebook to live-stream his attack on mosques in Christchurch, New Zealand (Isaac & Metz, 2019). Following this event, Schroepfer led Facebook’s efforts to increase its censoring technologies, using artificial intelligence (AI) to remove harmful or disturbing content before it can reach the platform’s users. Additionally, Chief Executive Officer of Facebook, Mark Zuckerberg was extremely disturbed about the incident and its tie to Facebook. In response to the incident, he stated, “‘Over a five- to 10-year period, we will have A.I. tools’ that can detect and remove hate speech” (Isaac & Metz, 2019).

While Facebook continued to use these technologies to eliminate content that may “cause harm,” the company used these tools only minimally to check the accuracy of information (Wells, 2019). Regarding political advertisements, Schroepfer stated that, “If this speech is going to cause … harm, I think most people think that’s a good place to start intervening” (Wells, 2019). Schroepfer was outspoken in his conviction that Facebook must eliminate content harmful to its users. However, he did not believe that the company should censor paid content based upon its level of accuracy (Wells, 2019). Because Facebook’s stance on censoring false information was notoriously vague, the impact of these filtering efforts remained somewhat unclear to the public.

The case study follows Facebook’s executive decision-making process and attempts to determine how to filter information on the company’s platform to ensure that unwanted content would not reach innocent users. Additionally, the case study evaluates how Facebook determined what constituted “harmful” content and what approaches the company employed to filter these posts. This case study is important because it provides a specific example of how this issue impacts both businesses and individuals on many levels: financially, psychologically, and socially.
Literature Review of Scholarly Articles

The literature review includes information from various scholarly journals and addresses the topic of information censorship on social media from different scholars’ perspectives. Furthermore, it examines scholars’ view about the impacts these platforms’ efforts have on the platforms’ users and on the companies themselves. The review analyzes each source to include all information that directly relates to the thesis topic. Additionally, the literature review provides an overview of the body of scholarly knowledge available on this topic to inform readers about the breadth of this discussion.

The published scholarly information available on this subject dates back to the creation of the United States (U.S.) Constitution in 1787 when Constitutional Convention delegates first addressed the issue of individuals’ rights to freedom of speech in the country. The scholars this review discusses differ in their opinions regarding the use of censorship and its ability to limit individual’s freedom of speech, but they do agree that some degree of freedom of speech is necessary to facilitate the unfettered exchange of ideas among individuals.

Against this background, the literature review next shifts its focus to evaluate what scholars believe constitutes misinformation, and health-related misinformation, specifically. These viewpoints differ according to how various scholars’ categorize and classify misinformation.

When considering how health-related misinformation spreads on social media, leading scholars found that three significant factors contribute to misinformation outbreaks: the human confirmation bias, information silos, and echo chamber effects. However, the scholars’ opinions regarding which the specific psychological influences further increase the spread of misinformation differed according to their analytic perspectives.
The literature review next considers the most effective means for companies to remove and correct misinformation on social media platforms. Many scholars agreed about the need for user-centered education programs to combat misinformation, but differed in terms of what education they believed would be most appropriate. Moreover, other leading scholars focused on ways to overcome the human confirmation bias and proposed various corrective measures to do so, including theory-based and information-based information corrections. Furthermore, in addressing additional factors that might affect misinformation corrections, the scholars identified various influences, including familiarity, information credibility, and source credibility.

Finally, the scholars offered insights on misinformation correction techniques specific to the COVID-19 pandemic. While all but one of the scholars emphasized the need for credibility and authority in information sources, they held varying views about how authorities should frame their information for social media users.

The literature review summarizes all of the scholarly articles and offers a conclusion specific to the COVID-19 pandemic. The review also recommends that scholars perform future research on the social media trends during misinformation outbreaks, the intentions of online misinformation content creators, and the impact of this misinformation on consumers and social media companies.

Instructor’s Manual with Teaching Notes

The instructor’s manual provides specific scholarly sources that instructors can draw upon when teaching this subject. Additionally, the instructor’s manual offers a summary of the case, as well as suggestions on how to appropriately teach the case study in a classroom setting. Information censoring on social media platforms has a significant impact on businesses and individuals. This discussion topic is appropriate for multiple business management, marketing, and finance courses because students will be able to evaluate this topic from multiple perspectives with various focuses.
Furthermore, because of the multi-faceted nature of the case, students will also be able to cross-reference the three disciplines to show how the topic interrelates all aspects of business. For example, students could analyze how the marketers on these social media platforms feel the financial impact of managerial decisions to censor some online content. In this regard, an ideal course for this subject would be a social media marketing course that analyzes the management of these platforms and examines the financial planning in their business models.

Additionally, students could discuss whether filtering information on social media platforms is necessary and explore the ethical consequences of this policy. Students could also examine what criteria Facebook used to identify harmful content and whether the company considered false information or lack of information harmful, as well.

After reading the case study, students should be familiar with Mike Schroepfer’s dilemma and be able to answer the following questions that appear in the instructor’s manual including: What is Facebook currently doing to proactively and retroactively reduce online misinformation spreads on its platform? How does misinformation negatively impact Facebook’s brand image, and how does filtering false or harmful information help Facebook's brand equity? What risks does Facebook face by limiting users’ rights to complete freedom of speech on the platform? How did the COVID-19 pandemic influence the spread of online misinformation throughout the social media industry and how did the pandemic impact the responses from various online companies?

**Conclusion of Thesis**

This thesis approaches the topic of misinformation on social media using a case study that tells the story of Facebook’s responses to the issue, a literature review that examines various scholars’ viewpoints on the issue, and an instructor’s manual that highlights the learnings objectives and teaching strategies of the topic. Together, these three sections present a cohesive overview of a problem that future researchers should explore in greater depth.
REFERENCES


Social Media Censorship: The Impact Filtering Information Has on the Well-Being of Users and Online Businesses

“In order to innovate, you have to experiment.” -Mike Schroepfer, Facebook, Inc. Chief Technology Officer

Facebook’s History

The Face Behind Facebook

Mark Zuckerberg, creator and initial CEO of Facebook, was born in White Plains, New York, on May 14, 1984, into a well-educated family (Tarver, 2020). Zuckerberg developed an early interest in computer science, and at the age of 12, he created a messaging program called “Zucknet” for his father’s dental office (Tarver, 2020). After seeing his computer programming interest, Zuckerberg’s parents hired a computer tutor to meet with Zuckerberg weekly (Biography.com Editors, 2019). His tutor, David Newman, later admitted that keeping up with Zuckerberg, a computer programming prodigy, was challenging (Biography.com Editors, 2019).

During this time, Zuckerberg also began taking graduate courses at Mercy College to further his knowledge (Biography.com Editors, 2019). Zuckerberg attended a preparatory high school, Phillips Exeter Academy, in New Hampshire (Biography.com Editors, 2019). While still in high school, he created a music software program called Synapse (Biography.com Editors, 2019). Zuckerberg’s Synapse was similar to Pandora, a popular music application that emerged five years later (Van, 2016). This application gained many companies’ attention, including both America Online, Inc. (AOL) and Microsoft (Biography.com Editors, 2019). These companies offered to buy the software from Zuckerberg and hire him immediately, but he declined their proposals (Biography.com Editors, 2019).
After graduating from high school in 2002, Zuckerberg enrolled at Harvard University (Tarver, 2020). There, Zuckerberg quickly established a reputation as the premier software engineer at the university (Biography.com Editors, 2019).

The Early Days

During his time as an undergraduate student at Harvard University, Zuckerberg created a social media platform he initially called “The Facebook” (Boyd, 2019). The Facebook, however, was not Zuckerberg’s first popular website on the Harvard campus (Boyd, 2019). In fact, Zuckerberg’s first website launch at Harvard was a program he named “CourseMatch” (Biography.com Editors, 2019). This software allowed students to pick their courses based on their peers’ schedules (Biography.com Editors, 2019). In 2003, Zuckerberg launched his next project, “Facemash” (Biography.com Editors, 2019). This site prompted students to rate the attractiveness of their peers on the online website (Boyd, 2019). A few hours after Zuckerberg launched Facemash, users had already viewed photos on the site 22,000 times (Boyd, 2019). Despite the website’s popularity, because of the website’s inappropriate nature, Zuckerberg had to meet with the Harvard Administration Board (Boyd, 2019). Nevertheless, Facemash brought much attention to Zuckerberg in the Harvard community (Boyd, 2019).

In the wake of the Facemash website’s popularity, three Harvard students, Divya Narendra, Cameron Winklevoss, and Tyler Winklevoss, solicited Zuckerberg’s help for a proposed project (Biography.com Editors, 2019). These three students had the idea to create a social networking site called “Harvard Connect” that would act as an online network for Harvard University students (Biography.com Editors, 2019). While Zuckerberg initially agreed to assist with Harvard Connect, he soon left to develop his own social media platform, The Facebook (Biography.com Editors, 2019).

Zuckerberg enlisted his friends, Dustin Moskovitz, Chris Hughes, and Eduardo Saverin, to help develop the new platform (Biography.com Editors, 2019). They launched The Facebook in February of 2004 (Boyd, 2019). The site originally included individual profile pages where users could upload their photos and interests and send “friend requests” to their peers (Boyd, 2019). At first, The Facebook was only available to Harvard University students (Boyd, 2019). The website was wildly popular on campus,
with 50 percent of Harvard’s students signing up within the first month (Boyd, 2019). As The Facebook rapidly gained popularity, Zuckerberg withdrew from Harvard and moved to Palo Alto, California to pursue his newly-developed company (Boyd, 2019). By the end of 2004, The Facebook had spread to almost all college campuses in the United States and Canada and the platform had reached one million users (Biography.com Editors, 2019).

The Company’s Unprecedented Growth

After Accel Partners invested $12.5 million in Facebook in exchange for a 15 percent share of the company in April of 2005 (DealBook, 2012), Zuckerberg changed the company’s name to “Facebook” and purchased the “facebook.com” domain for $200,000 (Boyd, 2019). The rate of Facebook’s growth continued to be unprecedented, with membership increasing from 12 million in December of 2006 to 50 million by October of 2007 (Boyd, 2019). With this increased growth in membership, Facebook began expanding its website into an all-encompassing platform (Boyd, 2019).

In May of 2007, Facebook launched its Marketplace, a platform for users to post and sell their products and services (Boyd, 2019). The company also introduced the Facebook Application Developer platform that allowed third-party developers to design applications and games within the Facebook site (Boyd, 2019). During this expansion, Facebook also increased its appeal to businesses and had over 100,000 companies signed up for the platform by the end of 2007 (Boyd, 2019). The site further increased connectivity among users in April of 2008 when it introduced Facebook Chat, a feature that allowed users to send direct messages to each other (Boyd, 2019).

By December of 2009, Facebook was the most popular social media platform globally, with 350 million registered users (Boyd, 2019). Less than a year later, in November 2010, analysts valued Facebook at $41 billion, making Facebook the third-largest web company in the United States (Boyd, 2019). Moreover, when Facebook purchased Instagram in April of 2012 for $1 billion, the company significantly expanded its social media empire (Boyd, 2019). This exponential growth firmly established Facebook as a critical player in the developing social media industry.
Facebook’s Growing Community and Political Influence

From its beginnings, Facebook has had a long-standing, complicated relationship with the United States’ political system. Over the years, the social media website has wavered in its political involvement. Nevertheless, it has influenced U.S. elections by providing users with a platform to share their political viewpoints.

Facebook’s first high-profile involvement with U.S. politics was during the 2008 Presidential election. As Facebook continued to increase its presence as a multifunctional social media platform, the company had also become a dominant news source for its users. In the months leading up to the 2008 historic election, many candidates used the Facebook platform to reach American voters (Dutta & Fraser, 2008). As a result, this election propelled Facebook into the spotlight of the complicated United States political system (Dutta & Fraser, 2008). In fact, a study by the Pew Research Center found that 46 percent of Americans used the web, e-mail, or text messaging to access news about the 2008 Presidential election, present their political opinions, or encourage others to vote (Dutta & Fraser, 2008). President Barack Obama gained tremendous support from his over 2 million young followers on the social media platform, resulting in what many call the “Facebook Election” of 2008 (Dutta & Fraser, 2008). This election was crucial in shaping Facebook’s future path to becoming more than solely a social media company.

Furthermore, Facebook demonstrated the effectiveness of targeted advertisements in a case study the company released in 2011. This study examined how the “Vote NO on 8” campaign in Florida had defeated a proposition on the November 2010 ballot that would have permitted large kindergarten through twelfth grade class sizes in state public schools (Facebook, 2011).

This campaign used Facebook as a “market research tool” to poll different demographic groups of Florida voters (Facebook, 2011). In the month before the election, the campaign placed targeted advertisements opposing the ballot proposition towards specific demographics (Facebook, 2011). These targeted groups included people who had relevant “likes & interests,” such as politically-oriented Facebook pages, or “education & work” terms, such as “teacher” or “PTA” on their Facebook pages (Facebook, 2011).
The “Vote NO on 8” campaign was successful. Voters who had the most ad exposure via Facebook were 17 percent more likely to vote against the proposition than those who had the least exposure (Facebook, 2011). The 2010 “Vote NO on 8” campaign was the first statewide election initiative to win using online advertising as the primary form of communication with voters (Facebook, 2011).

Following this success, Facebook continued to use its power in politics with the 2012 U.S. Presidential election. A research study led by James Fowler in 2012 found that Facebook’s “I Voted” button increased voter turnout by a third of a million people, most of who were young individuals (Ferenstein, 2012). While Facebook hoped to increase voter turnout and did not explicitly state support for any political party, its campaign likely had the effect of increasing Democratic votes because compared to the general U.S. population, young voters are more inclined to support the Democratic Party (Madrigal, 2017). The success of Facebook’s “I Voted” button proved Facebook’s immense power to influence community outcomes, thus establishing Facebook as a powerful, all-encompassing platform (Madrigal, 2017).

**2016 Presidential Election**

The 2016 Presidential election between Hillary Rodham Clinton and Donald John Trump further highlighted Facebook’s impact on U.S. political campaign outcomes. Following the massive news coverage of the election, many observers criticized Facebook for circulating misleading information and being a source of deceptive news for the public (Boyd, 2019). A study by Guess, Nyhan, and Reifler in 2016 found that over one-quarter of adults had visited a fake news website that supported Clinton or Trump in the final weeks leading up to the 2016 election (Kurtzleben, 2018).

Additionally, Facebook reported that the Internet Research Agency, a Russian company with ties to the Russian government, had uploaded approximately 80,000 pieces of content and spent $100,000 on about 3,000 advertisements that reached an estimated 29 million users on the site between January 2015 and August 2017 (Isaac & Wakabayashi, 2017).
Facebook’s General Counsel, Colin Stretch, claimed that these posts were “an insidious attempt to drive people apart” (Isaac & Wakabayashi, 2017). Analysts believed that Russia, and other countries that created these fake accounts, spread these messages and conspiracy theories to alienate Americans from one another prior to the 2016 election (Frenkel & Barnes, 2020). Yet, despite accusations regarding Facebook’s political election interference, the company maintained that:

Voting is a core value of democracy and we believe that supporting civic participation is an important contribution we can make to the community. We as a company are neutral – we have not and will not use our products in a way that attempts to influence how people vote. (Madrigal, 2017)

Similarly, when addressing accusations about Facebook’s influence on the election, Mark Zuckerberg called the notion that Facebook affected the outcome of the 2016 Presidential election a “pretty crazy idea” (Isaac, 2016).

**Company’s Response to Misinformation**

Throughout its history, Facebook shifted its views on the balance between protecting users by filtering posts and advertisements on its platform and remaining a neutral platform where users could freely post. However, in 2015, users began to discuss the prevalence of misinformation on the site. In response, Facebook took strong measures to control the spread of misinformation on the platform. At that time, Facebook attempted to reduce misinformation on the platform by introducing a new feature that allowed users to flag a questionable article as a “false news story” (Boyd, 2019). If enough users reported an article, Facebook flagged the article as misinformation (Boyd, 2019). While this feature was somewhat effective, the potential harm of false news became especially pressing after a study from the Pew Research Center in July of 2015 found that 63 percent of American adults used Facebook to receive news (Boyd, 2019).

Additionally, during the 2016 Presidential election, Facebook users brought increased pressure on the company to stop the spread of false news. Consequently, Facebook announced that it would begin working with fact-checking organizations to verify the information posted on its site (Boyd, 2019). Facebook also altered its advertising policies to limit advertisements that promoted false information (Boyd, 2019).
Despite these changes, Facebook still had not taken proactive measures to eliminate the spread of misinformation on the platform. With the lack of strong results from these initiatives, coupled with users’ increasing attention on the issue of false information, in 2017, Facebook announced a three-step proactive approach to address the misinformation problem (Boyd, 2019). Facebook’s management stated that it hoped to stop the spread of misinformation through three main initiatives: by disrupting economic incentives for posting misleading information, helping people make informed decisions, and building new technology products to filter false information and prioritize accurate posts (Facebook, n.d.).

**Disrupting Economic Incentives**

As Facebook users began to question the preponderance of false information on the site, Facebook prioritized its efforts to remove the economic incentives that encouraged misleading and inaccurate information posts (Facebook, n.d.). Prior to this initiative to disincentivize misleading information, many unethical traffickers were posing as legitimate news sources. These opportunistic traffickers created misleading posts with shocking headlines to attract users to view their websites (Facebook, n.d.). These traffickers rapidly increased their advertising profitability by luring unknowing consumers into false “news” websites that contained little actual news, and mostly advertisements (Facebook, n.d.).

To reduce economic incentives to false “news” traffickers, Facebook implemented an updated algorithm in 2018 that prioritized posts from families and friends over those of advertisers or companies (Boyd, 2019). However, the new algorithm nearly eliminated organic reach from legitimate companies to users (Boyd, 2019). Consequently, many companies saw a decrease in their user reach (Boyd, 2019). These companies could sustain their prior reach levels only if they began paying for ads on Facebook (Boyd, 2019). Previously, when users “liked” a company, the company’s posts would appear on users’ Facebook “feeds” (Boyd, 2019). However, the new algorithm eliminated this “free advertising” for companies (Boyd, 2019).
Meanwhile, Facebook tightened its advertising regulations. The new policies imposed strict measures to prevent creators from purchasing advertisements for content that contained false information (Facebook, n.d.). Because Facebook was displaying fewer false news advertisements on the site than previously, this policy change reduced misinformation on the platform (Facebook, n.d.).

Additionally, Facebook increased its use of community and third-party fact-checking organizations to detect and eliminate content that contained misinformation (Facebook, n.d.). Facebook applied these fact-checking measures and employed artificial intelligence (AI) to help the response teams identify false information (Facebook, n.d.).

**Helping People Make Informed Decisions**

While Facebook aimed to completely eliminate false information on its site, the company also worked to educate users about how false news spreads, offer ways to detect misinformation, and provide steps to take when encountering this deceptive content (Facebook, n.d.).

Following this approach, Facebook launched the Facebook Journalism Project in 2017 to prevent fake news from influencing users (Boyd, 2019). This platform encouraged users to support local news outlets and provided training in both journalism and fake news detection (Boyd, 2019).

Additionally, the company instituted the News Integrity Initiative with a group of over 25 funders and participants, including technology industry leaders, academic institutions, non-profit organizations, and third-party entities (Facebook, n.d.). This initiative focused on helping people form informed opinions about news that appeared on social media sites (Facebook, n.d.). The News Integrity Initiative aimed to increase social media users’ news literacy, rebuild trust in journalism globally, and inform communities about the issue of misinformation (Facebook, n.d.).
Building New Products

In addition to increasing users’ awareness about misinformation, Facebook developed new technology products to detect and eliminate false news on its site. These new products were essential to reduce incorrect information on Facebook; the software automatically checked the validity of users’ Facebook posts (Facebook, n.d.). When describing the company’s role in fact-checking information, Facebook administrators stated, “We cannot become arbiters of truth ourselves – it’s not feasible given our scale, and it’s not our role” (Facebook, n.d.).

These new products improved users’ abilities to report misleading or inaccurate posts (Facebook, n.d.). With this improved technology, the Facebook community could more easily report a false news story than previously. These reports reduced the number of inaccurate posts that reached users’ feeds (Facebook, n.d.).

Additionally, Facebook introduced ranking improvements in its feed algorithm (Facebook, n.d.). Facebook improved its algorithm by using new technologies that tracked users’ activities to determine how to prioritize posts for other users (Facebook, n.d.). For example, if a significant number of users were unlikely to share an article after reading it, the algorithm concluded that the story might have been misleading in some way (Facebook, n.d.). By applying these assumptions, the algorithm reduced the prevalence of misleading articles on users’ feeds (Facebook, n.d.).

Facebook also began looking towards a new technological advancement to slow the online spread of misinformation: artificial intelligence (AI). When appearing before the U.S. Congress in 2018, Zuckerberg testified that Facebook was developing new systems to “identify certain classes of bad activity” and claimed that, “Over a five- to 10-year period, we will have AI tools” that can remove harmful content from the platform (Isaac & Metz, 2019).
Facebook’s use of AI tools included a method called Deep Entity Classification (DEC) that evaluated the activities, behaviors, and interactions of suspected fake accounts (Statt, 2020). DEC reviewed around 20,000 characteristics of users to evaluate each profile and determine its legitimacy (Statt, 2020). Facebook used this tool to eliminate fake accounts, including those sophisticated programmers had created to appear as legitimate accounts (Statt, 2020). In 2020, Facebook claimed that this technology reduced the volume of spam accounts on the platform by 27 percent by eliminating more than 6.5 billion false accounts during 2019 (Statt, 2020).

**Facebook’s Errors in Post Filtering**

Facebook struggled to accurately and appropriately limit misleading, harmful, and factually incorrect information on its platform. In its role as a vehicle of free speech and unfiltered opinions, Facebook had to ensure that it did not incorrectly censor its users, especially when its users were presenting their posts as opinion pieces. Consequently, the company faced challenges to create an unbiased method to remove misinformation and protect its users. This responsibility was especially difficult because both users and fact-checkers could interpret the validity of many users’ posts subjectively.

Even with its improved technology products and algorithms, Facebook, nevertheless, still made errors by unintentionally filtering and removing factually correct information. In one high-profile instance, on October 14, 2020, Facebook’s filtering technologies identified a *New York Post* article as possibly containing inaccurate information and limited its distribution on the platform (Bond, 2020). The article contained allegations against Hunter Biden, the son of Joseph (“Joe”) Biden (Bond, 2020). At the time, the elder Biden was running for President of the United States in a contentious campaign. The *New York Post* article revealed emails that concerned Hunter Biden’s business dealings in China, and critics of Joe Biden claimed that the emails compromised the presidential candidate (Feis, 2020).
Facebook had limited the reach of this story on the platform before outside fact-checkers had completed their reviews of the claims (Bond, 2020). Andy Stone, Facebook’s spokesman, explained that Facebook would sometimes limit the spread of a story in an abundance of caution to give fact-checkers time to evaluate its validity if the platform had detected “signals” that the story could be false (Bond, 2020). Despite Stone’s clarification, Facebook’s seemingly inconsistent and non-transparent methods for filtering information drew criticism from both users and experts, alike.

Company’s Focus on Artificial Intelligence Initiative with New Chief Technology Officer

Mike Schroepfer, Facebook’s Chief Technology Officer (CTO) since 2013, led the team that worked on developing Facebook’s technology initiatives. Schroepfer was born on February 1, 1975, in Delray Beach, Florida (Metz & Isaac, 2019). He attended Stanford University and received two degrees: Bachelor of Science and Master of Science in Computer Science (CNBC, 2012).

Schroepfer was already a veteran in the technology and social media industry before joining Facebook. At the beginning of his career, Schroepfer worked at various startups, including Puffin Designs and Reactivity, Inc. (“Michael Schroepfer,” n.d.). While working for Puffin Designs from 1997 to 1999, Schroepfer focused on a product called “Commotion,” a digital video effects package. Multiple major motion pictures used “Commotion” for their special effects, including Gladiator, Dr. Dolittle, and Star Wars: The Phantom Menace (“Michael Schroepfer,” n.d.).

Schroepfer then became a partner at Reactivity, Inc.’s Consulting Practice, which connected highly skilled technical talent to small companies and startups (“Michael Schroepfer,” n.d.). In 2000, Schroepfer co-founded CenterRun, a software company that Sun Microsystems acquired in 2003 (CNBC, 2012). After the acquisition, Schroepfer worked as CTO for Sun Microsystem’s N1 Data Center Automation Division from 2003 to 2004 (“Michael Schroepfer,” n.d.). Following this position, Schroepfer was Mozilla’s Vice President of Engineering from 2005 to 2008 (WallMine, 2020). During his time at Mozilla, Schroepfer aided in developing the Firefox web browser (CNBC, 2012).
Schroepfer joined Facebook as the Vice President of Engineering in 2008 ("Michael Schroepfer," n.d.). In this position, Schroepfer oversaw all of the company’s engineering, infrastructure, and mobile efforts (Isaac, 2013). With Schroepfer’s continued success in the industry, Laptop Magazine cited him as one of the 25 most influential people in mobile technology in 2008 (CNBC, 2012). Additionally, Schroepfer served on several boards: Ancestry.com’s Board of Directors from 2011 to 2012 (CNBC, 2012), Anita Borg Institute for Women and Technology’s Board of Trustees from 2011 to 2018, and Wealthfront Inc.’s Board of Directors from 2015 to 2017 ("Michael Schroepfer," n.d.).

After becoming Facebook’s CTO in 2013, Schroepfer led his team of more than 150 engineering specialists to develop the company’s technological advances (The Org), including implementing new AI tools (Metz & Isaac, 2019). Facebook used these AI tools internally to sort and erase millions of harmful or misleading posts (Metz & Isaac, 2019).

Schroepfer enhanced his team by hiring researchers whose concentrations were in neural networks (Metz & Isaac, 2019). Schroepfer hoped that recruiting candidates with PhDs from New York University, the University of London, and the Pierre and Marie Curie University in Paris would make Facebook’s AI tools comparable to those of Google (Metz & Isaac, 2019). Additionally, Schroepfer created the Applied Machine Learning Team that created real-world applications from Facebook’s AI technologies, including facial recognition, language translation, and augmented reality tools (Metz & Isaac, 2019).

The Paris terrorist attacks in late 2015 drastically shifted Facebook’s AI focus (Metz & Isaac, 2019). In these coordinated attacks in the city of Paris and the surrounding area, Islamic militants killed 130 people and wounded almost 500 others (Metz & Isaac, 2019). Victims of this attack and the outraged public criticized social media and internet platforms Google, Facebook, and Twitter for knowingly permitting the Islamic State group (“ISIS”) to recruit members, raise money, and spread anti-Western propaganda on the platforms (BBC News, 2016). After the attack, Mark Zuckerberg asked the Facebook’s Applied Machine Learning team to focus on combatting terrorism by identifying and eliminating inflammatory terrorist posts on the
platform (Metz & Isaac, 2019). In response, Schroepfer and his team began working to create AI technologies that could identify terrorist propaganda on Facebook and flag suspect posts for review by human curators (Metz & Isaac, 2019). Furthermore, after the 2016 U.S. Presidential election, which had contained numerous instances of false reports about both candidates on Facebook, Schroepfer’s team broadened this technology to identify unwanted content, including nudity and fake accounts (Metz & Isaac, 2019).

During his Facebook career, Schroepfer continued as an outspoken critic of Facebook’s role in spreading misinformation and harmful content. Schroepfer entered the public spotlight in 2019 when discussing the Christchurch, New Zealand shootings that the gunman had live-streamed on Facebook (Metz & Isaac, 2019). Referring to this event, Schroepfer stated, “It won’t be fixed tomorrow. But I do not want to have this conversation again six months from now. We can do a much, much better job of catching this” (Metz & Isaac, 2019). He continued, “We’re not where we’d like to be, so if people are criticizing us I get the criticism. I’m frustrated as anyone when we make a mistake or something happens” (Metz & Isaac, 2019). Additionally, when explaining the future of AI technologies, Schroepfer stated, “I don’t foresee a future anytime soon where we don’t need people involved in the loop, because these are fundamentally people issues, and deciding what’s hate speech and what’s misinformation is a human endeavor” (McCracken, 2019).

**Facebook’s Fight Against False News During the COVID-19 Pandemic**

Facebook’s fight against misinformation became increasingly difficult with the outbreak of a global pandemic in 2020. Scientists first identified COVID-19, or Coronavirus disease, in Wuhan, China (Georgia Department of Health, n.d.). The novel virus quickly spread globally, inciting panic among individuals and resulting in unprecedented lockdowns in most countries. Amidst the global pandemic that rapidly restricted nearly all aspects of daily life, many individuals had concerns, doubts, and unanswered questions about their health and safety. The scientific community’s initial lack of knowledge about this virus caused many people to look to social media for answers.
On Facebook, users began posting fictitious “cures” for COVID-19 as well as conspiracy theories about the virus’s alleged origins (Alpert, 2020). A report by Avaaz, a global human rights group, found that health misinformation on Facebook peaked during the pandemic, reaching its one-year high in April of 2020 (Dwoskin, 2020). Additionally, the report estimated that in the month of April, Facebook users viewed the top ten websites that promoted health misinformation four times more than websites of ten leading health institutions (Dwoskin, 2020). These reputable institutions included the World Health Organization (WHO) and the Center for Disease Control and Prevention (CDC) (Dwoskin, 2020).

Many of the sites that promoted false and misleading claims about the COVID-19 pandemic had millions of followers on Facebook (Dwoskin, 2020). In one example of inaccurate information, users spread posts stating that high volumes of Vitamin C and silver particles could cure the virus (Alpert, 2020). Additionally, users shared a viral film called “Plandemic” millions of times on Facebook prior to its removal (Dwoskin, 2020). The video claimed that wearing a mask could cause people to develop COVID-19 (Dwoskin, 2020). In another instance, more than 160 million Facebook users viewed and more than 6 million users liked or commented on an article that falsely claimed that the American Medical Association was “encouraging” doctors to over-report COVID-19 deaths (Dwoskin, 2020). These false claims of cures for COVID-19 and conspiracies for its causes mislead vulnerable Facebook users. This widespread misinformation created further uncertainty about the virus, its prevention, and its potential cures.

**Efforts to Resolve COVID-19 Misinformation**

In response to the misinformation crisis, Zuckerberg stated, “Through this crisis, one of my top priorities is making sure that you see accurate and authoritative information across all of our apps” (Lewis, 2020). As the COVID-19 pandemic intensified, Facebook quickly acted to address the dangerous spread of misinformation.
In an interview regarding Facebook’s response to widespread misinformation on its platform, Nick Clegg, Vice President of Global Affairs and Communication at Facebook, stated that:

One of the most important antidotes to misinformation is good information. We certainly have worked very closely with the [Centers for Disease Control and Prevention] and others to make sure that we are constantly providing a source of reliable, authoritative, credible information. And we’re proud of the work of that. Separately, of course, we remove content related to vaccines or anything else where it poses an impending and immediate real world harm. (USA Today Editorial Board, 2020)

The company worked to eliminate all forms of misinformation, including both inaccurate and harmful content.

Facebook stated in its Community Standards Enforcement Report released on August 11, 2020, that the company had removed over 7 million posts during Q2 of 2020 that contained false information that might be harmful to users (Lerman, 2020). These posts included claims that social distancing was not effective and that non-scientifically proven measures could cure individuals with COVID-19 (Lerman, 2020). Additionally, Facebook piloted a program that notified users if they had liked, reacted to, or commented on posts that contained false claims about COVID-19 (Lewis, 2020). These alerts encouraged users to visit the World Health Organization (WHO) website to read why Facebook had flagged the information as false (Lewis, 2020). Facebook spokesman, Andy Stone, stated that “From April to June, [Facebook] applied warning labels to 98 million pieces of COVID-19 misinformation...and directed over 2 billion people to resources from health authorities” (Dwoskin, 2020).

Furthermore, Facebook had to rely more on automated technologies to review content on the platform during April, May, and June than it had previously because fewer content reviewers were available because of the COVID-19 pandemic restrictions (Rodriguez, 2020). With technological improvements, however, Facebook was able to significantly increase its proactive detection rate from Q1 to Q2 of 2020 (Rosen, 2020). For example, the proactive detection rate for hate speech rose from 89 percent to 95 percent during this period (Rosen, 2020). Facebook stated that it would continue to use
“technology to prioritize the review of content that has the potential to cause the most harm” (Rosen, 2020). The company explained that it planned to continue to pair proactive technology with individual content reviewers in the future to most effectively eliminate harmful or misleading content (Rosen, 2020).

Additionally, Facebook reported that its appeals for misleading content reports were significantly lower than in previous quarters because Facebook was not permitting this option as frequently as it had previously (Rosen, 2020). Facebook prioritized removing harmful content and could not bear the additional risk of wasting valuable time to review appeals (Rosen, 2020).

While Facebook was proactive in identifying misinformation related to the COVID-19 pandemic, its responses varied widely. In response to the crisis, the magnitude of Facebook’s actions against misinformation depended on the perceived harmfulness of the content (Dwoskin, 2020). For instance, for extremely harmful content, such as posts that asserted that drinking bleach could kill COVID-19, Facebook removed these posts immediately (Dwoskin, 2020).

However, Facebook did not delete other posts that contained misinformation that users had phrased as opinion pieces, including false claims against vaccinations (Dwoskin, 2020). Instead, Facebook attempted to minimize the number of users these posts reached on the site (Dwoskin, 2020). For example, when asked about private “anti-vaxxer” groups, Nick Clegg stated that:

We don’t allow them to advertise or in any way try to attract people to the group. I mean, we’re actually at the moment actively looking to see whether we need to tighten up those provisions further. But it would be an extraordinary thing for us to act against debate where people cast aspersions on vaccines. (USA Today Editorial Board, 2020)

As Clegg’s statement indicated, Facebook had not yet established a definitive stance on what would be the most appropriate action when addressing posts that contained users’ potentially harmful points of view.
Facebook’s Dilemma Establishing Uniform Information Filtering Criteria

Mike Schroepfer and his team at Facebook faced the challenging dilemma of how to remain a community that facilitated free speech and user creativity while also protecting its users from inadvertently being exposed to misleading and harmful content.

However, in the process, Facebook’s business model was centered around providing a “free” service to users. Users could see questionable advertising content embedded in seemingly innocent posts from other users. Facebook’s financial success depended on the company’s profits from advertisements, and Facebook’s user activity and user reach influenced advertisers’ willingness to pay. Nevertheless, Facebook had been transparent about stopping the spread of misinformation on its platform, even though this action reduced the site’s overall user activity.

Facebook’s strategy to reduce misinformation on its platform raises several questions: Will filtering false and harmful information ultimately be profitable for Facebook, or is filtering a corporate social responsibility cost to establish credibility in the brand? How can Schroepfer determine what posts constitute “misinformation” when users could interpret the information subjectively, such as with opinion pieces? Should Schroepfer treat misinformation during the COVID-19 pandemic differently than previous cases of health-related misinformation? How should Schroepfer balance the risk of unintentionally removing accurate information on the one hand against the risk of failing to identify inaccurate information on the other hand? Should Facebook track users’ activity to proactively detect and remove false information?
APPENDIX

Exhibit 1

Russian-Purchased False Advertisements in 2016
Exhibit 2

2016 Facebook Adjusted Close Stock Price (USD)

Exhibit 3

2017 Facebook Adjusted Close Stock Price (USD)
Exhibit 4

2020 Facebook Adjusted Close Stock Price (USD)

Exhibit 5

Facebook Monthly Active Users Worldwide (Millions)
Exhibit 6

Facebook Revenue from 2014-2019 (in millions)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Advertising</td>
<td>69,655</td>
<td>55,013</td>
<td>39,942</td>
<td>26,885</td>
<td>17,079</td>
<td>11,492</td>
</tr>
<tr>
<td>Other Revenue</td>
<td>1,042</td>
<td>825</td>
<td>711</td>
<td>753</td>
<td>849</td>
<td>974</td>
</tr>
<tr>
<td>Total Revenue</td>
<td>70,697</td>
<td>55,838</td>
<td>40,653</td>
<td>27,638</td>
<td>17,928</td>
<td>12,466</td>
</tr>
<tr>
<td>Total Revenue % Change from Previous Year</td>
<td>26.6%</td>
<td>37.4%</td>
<td>47.1%</td>
<td>54.2%</td>
<td>43.8%</td>
<td>58.4%</td>
</tr>
</tbody>
</table>

Exhibit 7

Confounding Variables in Facebook’s 2020 Stock Price

<table>
<thead>
<tr>
<th>Confounding Variables in Facebook’s 2020 Stock Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>Increase in users’ and advertisers’ social media use during COVID-19 pandemic</td>
</tr>
<tr>
<td>Increase in competition in social media industry</td>
</tr>
<tr>
<td>V-shaped recovery in financial market following COVID-19 outbreak</td>
</tr>
<tr>
<td>Facebook ad boycott by over 400 brands</td>
</tr>
<tr>
<td>Facebook addition of new e-commerce feature: Shops</td>
</tr>
</tbody>
</table>
REFERENCES


Social Media Censorship: The Impact Filtering Information Has on the Well-Being of Users and Online Businesses

Research Question: When health-related misinformation spreads on social media platforms, what are the appropriate responses for these platforms to reduce users’ exposure to false and misleading information, minimize potential adverse effects, and safeguard these platforms as conduits for free speech?

Introduction

The prevalence of health-related misinformation has been a long-standing dilemma, both within individual communities and globally. However, the spread of misinformation has accelerated with the popularity of social media platforms and the online services community. While social media platforms have increased the accessibility of news and information to users globally, they have also contributed to the creation and distribution of false, user-generated content. Internet users create and spread misinformation with various conscious or subconscious intentions. Regardless of the motivation, the spread of misinformation, especially when health-related, can be detrimental to both users and businesses. To eliminate misinformation and prevent its future spread, social media companies and individuals must be vigilant in their evaluations and responses to new information.

This literature review will use scholarly articles to examine why internet users create and spread misinformation and will explore the most effective approaches for companies to eliminate and prevent misinformation on their platforms. The paper will evaluate the points of parity and points of difference among the scholars’ conclusions.
This literature review will first examine the history of freedom of speech in the United States and the limitations on this First Amendment right. The paper will discuss the United States Bill of Rights and Section 230 of The Communications Decency Act (CDA) of 1996. Section 230 protects online service providers from being liable for user-generated content posted on their platforms (Congressional Research Service).

The paper will next evaluate the effectiveness of censorship policies on reducing misinformation and these policies’ potential to infringe on individuals’ right to freedom of speech. This discussion leads to the topic of how scholars define misinformation, including if users’ intents are necessary considerations in classifying inaccurate information.

The literature review will also compare censorship in the social media industry to that of other industries. This evaluation will use the advertising restrictions in the prescription drugs industry as an example of how businesses use third-party regulators to control and reduce misinformation. This analysis will also include a contradictory viewpoint that asserts that social media is overbearing in its control of users’ exposures to selected information.

The literature review will then identify the critical contributing factors to the spread of misinformation, including individuals’ psychological influences, their perceptions of the importance of the information, and the ambiguity of relevant evidence on any particular topic of interest. This assessment will include the psychological processes and biases that humans experience when receiving and evaluating information, including the human confirmation bias, information silos and echo chamber effects, and the human need for social approval and reciprocity.

Next, the literature review will examine the impacts of misinformation on both individuals and social media organizations. This section will highlight the distrust that the general public has regarding the healthcare community. It will also investigate the potential problems with social media platforms’ lack of corporate social responsibility (CSR) initiatives and misinformation censoring policies to protect their users’ safety.
The review will also examine scholars’ beliefs about the most effective means to correct and or remove misinformation on social media. These solutions include scholars’ recommendations about how to educate users to identify and refute misinformation, and how users can overcome human psychological biases. Additionally, the paper will include scholars’ proposals about the most effective forms of information corrections and their evaluations of the importance of familiarity and credibility in correcting misinformation.

Finally, the paper will discuss various strategies of misinformation correction campaigns for the 2021 COVID-19 vaccination initiatives in the United States. This discussion will include scholars’ opinions about the causes behind misinformation during the COVID-19 pandemic and their suggestions about how to resolve concerns and eliminate misperceptions about the COVID-19 vaccines and vaccinations. The methods scholars propose to increase COVID-19 vaccination uptake include emphasizing community well-being, using credible information sources, and explaining the benefits of the vaccines and the costs of not receiving the vaccines.

The literature review will conclude with recommendations for further research, including improving tracking and collecting data on online trends during misinformation outbreaks and analyzing the detailed impacts of misinformation on the profitability, customer lifetime value (CLV), and brand equity of social media companies.

**History of Free Speech in the United States**

The Bill of Rights, which comprises the first ten amendments of the United States Constitution, contains a guarantee to intellectual freedom as its First Amendment (Harer & Harrell, 2002). The First Amendment includes four basic freedoms for U.S. citizens: freedom of religion, freedom of speech, freedom of the press, and the right to assembly (Harer & Harrell, 2002). Through various court cases over time, individuals and groups have asked the courts to interpret the Constitution. In response, the courts have identified five categories of limitations on free speech (Harer & Harrell, 2002). These categories include libel, slander, dissent, obscenity, and limitations on commercial speech (Harer & Harrell, 2002).
The rapid and dispersed spread of information through new technologies has complicated individuals’ rights to freedom of speech. Following the development of online services, internet providers became concerned about the liabilities of having false or injurious user-generated content on their platforms. In response to these and other concerns, the Communications Decency Act of 1996 (CDA) added Section 230 to the Communications Act of 1934 (Congressional Research Service, n.d.). Section 230 of the CDA protects online service providers from “legal liability stemming from content created by the users of their services, with some exceptions” (Congressional Research Service, n.d.). This Section allows “interactive computer service” providers, such as Facebook, Twitter, and Google, to publish users’ content without reviewing it for any “criminality or other potential legal issues” (Congressional Research Service, n.d.).

According to Section 230(c)(1), “No provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider” (Congressional Research Service, n.d.). While Section 230 of the CDA protects online service providers from many liabilities related to content that users post on internet platforms, the Section has some exceptions to this liability protection (Congressional Research Service, n.d.). Section 230’s coverage does not include user-generated content that violates Federal criminal law, the Electronic Communications Privacy Act of 1986, state laws similar to the Electronic Communications Privacy Act of 1986, or that infringes on intellectual property rights (Congressional Research Service, n.d.).

Some lawmakers, and the general public, have expressed concern over the extensive protections guaranteed under Section 230 and question its ability to protect online service providers from liability for the content posted on their websites (Congressional Research Service, n.d.). In response to these concerns, some Congressional members have introduced bills to address these concerns with the goal to hold online service providers accountable for sex trafficking-related civil and state criminal actions, and to amend the Federal criminal sex trafficking statute (Congressional Research Service, n.d.).
Evaluation of Internet-Content Censorship Policies

Henry Reichman, an attorney and former assistant director for the Office for Intellectual Freedom, a division of the American Library Association, defined censorship as, “The removal, suppression or restricted circulation of literary, artistic, or educational materials of images, ideas, and information on the grounds that these are morally or otherwise objectionable in light of standards applied by the censor” (Harer & Harrell, 2002).

Harer and Harrell (2002) identified four categories of complaints that Reichman asserted lead to some variation of censorship, including benign complaints, radical complaints, consensus complaints, and non-consensus complaints (Harer & Harrell, 2002). According to these scholars, a benign complaint refers to “a remark or statement that is not accompanied by a request for action” (Harer & Harrell, 2002). In contrast, a radical complaint is “an act of censorship that is taken by a group or person who objects to a creative work” (Harer & Harrell, 2002). Additionally, a consensus complaint occurs “when the authority to whom the complaint is made agrees with the person or group that complains,” and a non-consensus complaint is “when the person who complains and the authority to whom he or she complains disagree” (Harer & Harrell, 2002).

Harer and Harrell (2002) continued to discuss scholars’ views about the limitations on individuals’ freedom of speech and freedom of expression by referencing John Stuart Mill, a nineteenth-century philosopher. Mill argued that people could replace false ideas with true ideas when a social exchange of ideas occurred within the marketplace of ideas (Harer & Harrell, 2002). Mill claimed that if individuals lost their rights to express an opinion, everyone would suffer, including both those who agreed and those who disagreed with the opinion, because individuals would have no opportunity to trade false ideas for true ones (Harer & Harrell, 2002).

Siebert et al. (1984) agreed with Mill in his evaluation of censorship as inhibiting a necessary flow of information among individuals (Soley, 2002). They discussed four different topologies for organizing the world’s mass media (Soley, 2002). One of these organizational structures was what the scholars referred to as the “social responsibility system” (Soley, 2002). Under this system, the media would operate under the assumption that freedom of speech is “a necessary implement for the attainment of a desired goal”
Additionally, the social responsibility system stated that the press would be “obligated to be responsible to society” (Soley, 2002) by “enlightening the public about governmental affairs, serving as a forum for debate about public issues, and ‘maintaining its own financial self-sufficiency so as to be free from the pressures of special interests’” (Soley, 2002).

Siebert et al. (1984) further explained the problematic nature of media mergers that have “placed the mass media in fewer and fewer corporate hands,” resulting in large corporations that must meet earnings forecasts to appease shareholders (Soley, 2002). As a result, news outlets often prioritize profits over accuracy and increasingly rely on advertisers as their primary revenue sources (Soley, 2002).

Harer and Harrell (2002) also cited the contention of Thomas Emerson, Yale law professor and constitutional scholar (Harer & Harrell, 2002). Emerson stated that a distinction must exist between expression and action (Harer & Harrell, 2002). He believed that speech could lead to actions that may or may not be “objectionable,” so society must condemn actions that destroy public order (Harer & Harrell, 2002). However, Emerson believed the First Amendment must protect individuals’ rights for self-fulfillment, truth, participation in decision making, and balance between stability and change (Harer & Harrell, 2002). Emerson further asserted that when speech contributes to any of these values protected under the First Amendment, it must be protected, even in cases with an ulterior link between the speech and a subsequent destructive action (Harer & Harrell, 2002).

While Niemiec (2020) agreed with Emerson in his assessment that free speech must remain protected under the First Amendment, the scholar offered an additional criticism that censoring only information that directs action could be challenging (Niemiec, 2020). Niemiec (2020) stated that censorship is not a new concept, and social media platforms regularly and often remove “objectionable content” that violates their companies’ policies (Niemiec, 2020). He believed that the problem with categorizing information as “hate speech,” “glorification of violence,” or “harmful and dangerous content” is that these categories are vague and often interpreted subjectively (Niemiec, 2020).
Definition of Misinformation

To prevent the spread of “misinformation,” social media platforms must first appropriately define what they consider to be “misinformation.” Scholars vary in their definitions of “misinformation.” For example, Schoch-Spana et al. (2020) defined “misinformation” as “information that differs from expert consensus at the time it is shared” (Schoch-Spana et al., 2020). Meanwhile, Chou et al. (2018) described “health misinformation” as “a health-related claim of fact that is currently false due to a lack of scientific evidence” (Chou et al., 2018).

Additionally, Chou et al. (2018) considered all false health-related claims to be “health misinformation” and did not consider the intent or means of spread of the information in their assessment (Chou et al., 2018). However, Wang et al. (2019) made distinctions between various forms of misinformation and misleading information (Wang et al., 2019). Wang et al. (2019) claimed that “fake news” differs from other forms of misleading information in its intent and mode of spread (Wang et al., 2019). The scholars defined “misinformation” as “information that is inadvertently false and is shared without intent to cause harm” in contrast to their definition of “disinformation” as including “false information knowingly being created and shared to cause harm” (Wang et al., 2019).

Lazer et al. (2018) defined the commonly used term, “fake news,” as “fabricated information that mimics news media content” (Wang et al., 2019). However, Wang et al. (2019) noted that this definition of “fake news” did not fully encompass the complexity of the phenomenon that could also include “both satire and information created deliberately to mislead as a means to achieve a political or other goal” (Wang et al., 2019).
Review of Third-Party Regulators of Misinformation

While the United States Food and Drug Administration (FDA) had measures in place to prevent the spread of health-related misinformation about prescription medications, social media platforms did not have an equivalent regulator. Misinformation spread on social media rapidly and uncontrollably because the industry lacked clear guidelines and restrictions to hold the individual platforms accountable for the content they permitted and distributed on their websites.

United States FDA restrictions limited the spread of misinformation in relation to drug products (American Family Physician, 2010). In May of 2010, the United States FDA announced the launch of its Bad Ad Program (American Family Physician, 2010) to “prevent the misleading or inaccurate promotion of prescription medications” (American Family Physician, 2010). According to the FDA, misleading or inaccurate information about prescription medications involves “omitting or downplaying the risks associated with the product, overstating a product’s effectiveness, promoting off-label uses of products, and making misleading drug comparisons” (American Family Physician, 2010). The Bad Ad Program encouraged physicians to report potential violations of the FDA’s advertising and marketing regulations (American Family Physician, 2010). However, no similar program existed to regulate social media platforms’ spread of misleading or inaccurate claims regarding health-related topics.

Niemiec (2020), when addressing the lack of third-party organization regulators for social media websites’ content, criticized the often-accepted perspective that social media platforms are private companies and thus, have the right to decide what content they choose to censor (Niemiec, 2020). While the scholar agreed with the platforms’ rights to refrain from censoring content, he believed that the reason was not connected to the platforms’ private company statuses, but rather, rested on the platforms’ contentions that they were communication networks where users can freely express their opinions and views (Niemiec, 2020). Additionally, Niemiec (2020) noted that “A few big tech companies currently dominate social media services which also serve as a source of news to many users” (Niemiec, 2020).
In fact, a 2020 survey by the Reuters Institute for the Study of Journalism found that 36 percent of participants from 12 countries used Facebook for weekly news (Niemiec, 2020). Niemiec (2020) explained that these social media companies have a large influence over the content they disseminate to users (Niemiec, 2020). Niemiec (2020) believed that by censoring content based on vague guidelines, these platforms are effectively able to shape the viewpoints of many of their users by limiting the information they share (Niemiec, 2020). While Niemiec (2020) did not believe that social media platforms should have the right to censor content, the scholar also posed concerns about the use of third-party authoritative sources to regulate user content (Niemiec, 2020).

In response to misinformation spreading on social media during the COVID-19 pandemic in 2020, companies, including Facebook, Twitter, and YouTube, issued a statement stating that they were “jointly combating fraud and misinformation about the virus” (Niemiec, 2020). However, Niemiec (2020) explained that these companies all mentioned the World Health Organization (WHO) as their primary authoritative source (Niemiec, 2020). Niemiec (2020) emphasized that while some might claim that the WHO is a reliable organization, the organization may still make mistakes and might have ulterior motives when handling pandemics, especially in terms of the potential influence from pharmaceutical companies (Niemiec, 2020).

**Spread of Health-Related Misinformation**

The spread of misinformation is not a new problem and has been a prevalent concern at least since the beginning days of printing (Wang et al., 2019). However, the emergence of the internet has created a critical shift in the rate and extent of the spread of misinformation, especially in the fields of health and medicine (Wang et al., 2019). While scholars recognize the prevalence of health-related misinformation on social media, they greatly differ in their proposed causes for its widespread and rapid dissemination.

Vijaykumar et al. (2018) specified that social media platforms have the ability to shape users’ perceptions and behavioral responses to various public health emergencies (Vijaykumar et al., 2018). The scholars proposed that when addressing infection diseases crises, social media platforms are important considerations for healthcare authorities because of a platform’s “ability to amplify health issues through online information
diffusion” (Vijaykumar et al., 2018). Vijaykumar et al. (2018) stated that information on social media is incredibly malleable because individuals and institutions can direct mass attention to particular issues and stories, regardless of their sources’ credibility (Vijaykumar et al., 2018). They suggested that health and risk information is spread and amplified during an emerging infectious disease outbreak (EIDO) through dissemination, diffusion, and the exchange among individuals and institutions across multiple communication channels (Vijaykumar et al., 2018).

Chou et al. (2018) differed from Vijaykumar et al. (2018) in their explanation for why misinformation spreads on social media platforms. Chou et al. (2018) stated that misinformation spread on social media is the consequence of individuals’ lack of trust in institutions (Chou et al., 2018). In fact, the scholars pointed to a 2016 Gallup report that found that only 36 percent of individuals in the United States had sufficient confidence in the medical system (Chou et al., 2018). Additionally, 20 percent of individuals expressed skepticism about scientists (Chou et al., 2018).

Wang et al. (2019) had a third explanation for misinformation spread on online platforms. The scholars cited the “basic law of rumor” to explain the spread of misinformation (Wang et al., 2019). The “basic law of rumor” states that the amount of rumor spread will vary according to individuals’ perceptions of the importance of the subject multiplied by the ambiguity of the evidence relevant to the discussion topic (Wang et al., 2019). Wang et al. (2019) referred to Allport and Postman’s (1947) research that suggested that users’ evaluations of the ambiguity of a message may be influenced when they receive conflicting information from comparably-credible sources (Wang et al., 2019). However, Wang et al. (2019) asserted that assessing an information’s credibility is especially challenging on social media because users self-publish posts and face no tests of verification or accountability (Wang et al., 2019).

Chou et al. (2018) claimed that information silos and echo chamber effects further encourage the spread of health-related misinformation (Chou et al., 2018). In these information silos, each of which contains a separate collection of similar information, users are not exposed to differing viewpoints; consequently, they face the increased risk of spreading misinformation (Chou et al., 2018). In fact, Chou et al. (2018) maintained that individuals’ information exposure on social media depends largely on their previous
beliefs and biases (Chou et al., 2018). Furthermore, social media platforms have become experts at tailoring posts and advertisements on users’ feeds to match their interests and personal identities (Chou et al., 2018). This perpetual cycle of information sharing leads to information silos (Chou et al., 2018).

Niemiec (2020) expanded on the psychological drivers of misinformation by stating that the human psychology on which social media sites depend, such as the need for social approval, reciprocity, and novelty seeking, might contribute to the spread of misinformation (Niemiec, 2020). Niemiec believed that if social media users are aware of the business models of social media companies as well as the psychological mechanisms that influence why people use social media, then they may be able to avoid being manipulated by misleading content on social media platforms (Niemiec, 2020).

Impacts of Misinformation on Individuals and Companies

Individual Effects

Chou et al. (2020) highlighted the varied consequences of misinformation on those who are exposed to it (Chou et al., 2020). One potential consequence that Chou et al. (2020) cited is that misinformation may create the impression among viewers that official sources do not have a unified consensus on a particular issue (Chou et al., 2020). According to the scholars, this misperception that officials have no unified conclusion or solution on an issue may lead individuals to feel that official sources of information are not credible or trustworthy (Chou et al., 2020). The effects of these viewpoints could be detrimental as individuals begin to ignore accurate health information, avoid health care, or make detrimental health decisions based on false claims by unofficial sources (Chou et al., 2020).

Company Effects

In addition to the effects of misinformation on individuals, Grygiel and Brown (2019) discuss the impact that misinformation has on social media platforms’ brand images and their corporate social responsibility (CSR) impacts (Grygiel & Brown, 2019). These scholars asserted that in order to effectively address misinformation on social media and create safe platforms for users, the companies themselves must find value in
reducing misinformation (Grygiel & Brown, 2019). The scholars suggested that social media companies must increase their CSR efforts to raise their standards of information reliability for users (Grygiel & Brown, 2019). Grygiel and Brown (2019) explained that social media platforms perform content moderation, the “screening of [user-generated content] posted to social media platforms,” to ensure that each website offers an “environment that is hospitable to advertising as advertisers have become increasingly concerned with what type of content is affiliated with their campaigns” (Grygiel & Brown, 2019). The scholars explained that often social media platforms screen content not only because it is “offensive and harmful,” but also because failing to do so would harm the companies’ brand images and limit the companies’ profitability (Grygiel & Brown, 2019).

While many companies have attempted to moderate the content on their platforms, Grygiel and Brown (2019) stated that no unified standards for content moderation are in place across the industry (Grygiel & Brown, 2019). The scholars suggested that social media-based activism has influenced corporate changes in the social media industry because this activism encourages companies to institute CSR initiatives (Grygiel & Brown, 2019). For example, following the 2018 Facebook Cambridge Analytica scandal, the hashtag, “#deletefacebook,” trended on Twitter and drove Facebook to change its policies to maintain a positive relationship with its users (Grygiel & Brown, 2019). Social media companies react to users’ concerns and demands to remove problematic content and misinformation from their platforms because if they do not do so, they risk losing these valuable customers (Grygiel & Brown, 2019).

Ways to Correct Misinformation

While many scholars will attest to the fact that the spread of misinformation on social media is a vital issue, they largely differ in their opinions about how to most effectively reduce and correct misinformation on social media platforms. Their suggestions reflect scholars’ wide varieties of perspectives, including focuses on demographic tracking and education, psychological biases, and misinformation correction methods.
One of the most critical challenges in reducing health misinformation on social media involves determining the point at which social media platforms should intervene in users’ misleading posts (Chou et al., 2018). Chou et al. (2018) believed that to determine when social media companies should take action to remove misinformation, they should assess both the reach and the health consequences that the misinformation will have on specific populations, including identifying those who may be most vulnerable to the adverse effects of misinformation (Chou et al., 2018).

The scholars proposed that social media platforms can evaluate how individuals evaluate and internalize a message by using physiological measurements, real-time behavioral data, linkage to medical records, marketing research, and mixed-method approaches (Chou et al., 2018). Observational studies are one useful method the scholars suggested to observe the impacts of misinformation on individuals (Chou et al., 2018). These studies may include eye tracking, functional magnetic resonance imaging, or cognitive interviews that assess how quickly individuals accept a claim prior to internalizing it (Chou et al., 2018).

When addressing how social media platforms can reduce online health misinformation, Chou et al. (2018) proposed that these platforms should invest in improving users’ health and science literacy to bolster their trust in the medical community (Chou et al., 2018). The scholars believed that this proactive, broad approach might be more beneficial than if social media platforms attempted to eliminate all individual accounts of misinformation (Chou et al., 2018). Nevertheless, Chou et al. (2018) recognized that social media platforms must also continue to develop and apply mechanisms to filter and validate the accuracy and credibility of information on their websites (Chou et al., 2018).

Similar to Chou et al. (2018), Niemiec (2020) asserted that the solutions to minimize the potential harms related to misinformation include educating users and raising the public’s awareness about the dangers of inaccurate health-related information (Niemiec, 2020). However, Niemiec (2020) believed that this education should focus on two knowledge areas: social media and related business models; and science reporting, scientific research, and its limitations (Niemiec, 2020). Niemiec (2020) asserted that
knowledge will help users make informed decisions about how they use social media platforms (Niemiec, 2020). Specifically, Niemiec (2020) suggested that this education should explain that social media companies base their business models on revenues from advertisements that the companies tailor to specific user preferences (Niemiec, 2020). He also noted that the profitability of these companies depend on high user engagements for extended periods of time (Niemiec, 2020).

Furthermore, Niemiec (2020) recommended that users should recognize the limited reliability of scientific reporting (Niemiec, 2020). The scholar explained that news outlets can present biased or misleading science news, either intentionally or unintentionally (Niemiec, 2020). Scientific reports can include mistakes or biases and can even reflect influence from commercial interests or political or moral views (Niemiec, 2020). Niemiec (2020) believed that if the public is aware of the uncertain reliability of scientific reporting, social media users will be cautious about accepting unverified scientific information that appears online (Niemiec, 2020). This directly contradicted the opinion of Chou et al. (2018). Whereas Niemiec (2020) emphasized the unreliability of scientific reporting, Chou et al. (2018) emphasized the necessity for trust in scientific information (Chou et al, 2018). Chou et al. (2018) believed that public officials and health authorities must educate the public on the reliability and credibility of the medical community in order to decrease reliance on misinformation (Chou et al., 2018).

**Overcoming Psychological Biases**

Chou et al. (2020) expanded on Chou et al.’s (2018) perspective that psychological factors drive users’ spread of misinformation on social media platforms (Chou et al., 2020). Specifically, the scholars cited the possibility of the human confirmation bias that would make offering corrective information ineffective (Chou et al., 2020). When humans experience confirmation bias, they seek and affirm information that confirms their original beliefs and opinions and reject or ignore information that contradicts these views (Chou et al., 2020). Chou et al. (2020) stated that contrary to using corrective information to change users’ incorrect opinions, interventions based on value affirmation may be more effective in persuading individuals with strong confirmation biases (Chou et al., 2020).
Unlike Chou et al. (2020), Bode and Vraga (2015) concluded that social media platforms and their respective algorithms have the ability to shape users’ attitudes and reduce misperceptions (Bode & Vraga, 2015). Bode and Vraga (2015) conducted an experiment in a web-based survey to study the role that social media may play in correcting misinformation (Bode & Vraga, 2015). In this study, they focused on Facebook’s use of suggested links related to users’ current activities on the platform (Bode & Vraga, 2015). The scholars exposed users to a post containing misinformation and then altered the related stories to either confirm, correct, or both confirm and correct the misinformation (Bode & Vraga, 2015). One misinformation piece they used asserted that genetically modified organisms (GMOs) cause illness (Bode & Vraga, 2015).

The scholars found that they could change participants’ attitudes related to GMOs by exposing individuals to corrective information within social media, thus reducing individuals’ misperceptions (Bode & Vraga, 2015). Bode and Vraga (2015) emphasized that the attitude changes were not due to the presence of heuristic cues from members of an individual’s social network, but instead, were the result of articles suggested by Facebook’s algorithm (Bode & Vraga, 2015).

One common perception is that internet engineers design social media algorithms that intentionally limit the pool of information to which social media users can assess. However, Bode and Vraga (2015) contended that these algorithms can have a positive effect when the information they allow contained suggestions that correct misinformation (Bode & Vraga, 2015). Even though individuals often give related news stories lower evaluations when they contradict pre-existing attitudes, the scholars found that these lower evaluations do not prevent these stories from changing individuals’ attitudes who believed the misperception (Bode & Vraga, 2015).
Evaluating Correction Types

Gesser-Edelsburg et al. (2018) established new criteria for categorizing misinformation corrections (Gesser-Edelsburg et al., 2018). In their experimental design, the scholars sorted health organizations’ corrections into two conditions according to the correction method (Gesser-Edelsburg et al., 2018). Condition 1 was “common information correction” while Condition 2 was “recommended (theory-based) information correction, mainly communicating information transparently and addressing the public’s concerns” (Gesser-Edelsburg et al., 2018).

In the study, Condition 1, or information correction, included the most common way in which health organizations approach correcting social media misinformation, by labeling misinformation as “myths” (Gesser-Edelsburg et al., 2018). Gesser-Edelsburg et al. (2018) explained that studies have found this form of correction to be ineffective because labeling the information as myths allows individuals to remember the information, even though it is untrue (Gesser-Edelsburg et al., 2018). Additionally, the scholars asserted that individuals will not accept an information correction based purely on a classification change and without scientific evidence (Gesser-Edelsburg et al., 2018). Instead, Gesser-Edelsburg et al. (2018) hypothesized that Condition 2, or theory-based information correction, would be an effective correction method for misinformation (Gesser-Edelsburg et al., 2018).

According to the Gesser-Edelsburg et al. (2018), theory-based information correction involves health organization providing transparent information that addresses the public’s emotional concerns (Gesser-Edelsburg et al., 2018). The scholars believed that theory-based information correction would be more effective than common information correction (“myth-labeling”) because of the human confirmation bias (Gesser-Edelsburg et al., 2018). Gesser-Edelsburg et al. (2018) stated that:

People often tend to reject information corrections that contradict their attitudes, share content that is consistent with their own narratives while ignoring the rest, and allow the mistaken information to continue to effect [sic] their conclusions even when they bluntly admit the information is incorrect. (Gesser-Edelsburg et al., 2018)
However, the scholars stated that individuals have the opportunity to overcome the bias, citing other studies that proved that the correction of misinformation by organizations has a “positive effect on the public’s decision making” (Gesser-Edelsburg et al., 2018).

Gesser-Edelsburg et al. (2018) analyzed subjects’ responses to both types of information correction conditions when the surveyed participants viewed vaccination misinformation on Facebook (Gesser-Edelsburg et al., 2018). The researchers found that participants showed a significant decrease in their ratings of reliability between the two information correction conditions (Gesser-Edelsburg et al., 2018). Subjects rated the corrections in Condition 2 as having significantly higher average reliability levels than the corrections in Condition 1 (Gesser-Edelsburg et al., 2018). Additionally, subjects rated their average satisfaction levels as significantly higher in Condition 2 compared to Condition 1 (Gesser-Edelsburg et al., 2018). These results emphasized the importance of organizations correcting misinformation transparently (Gesser-Edelsburg et al., 2018). Furthermore, Gesser-Edelsburg et al.’s findings underscored the effectiveness of including emotional appeals to viewers through theory-based information corrections when social media platforms correct misinformation (Gesser-Edelsburg et al., 2018).

Similar to Gesser-Edelsburg et al. (2018), Vraga et al. (2020) explored the effectiveness of different intervention techniques on misinformation correction (Vraga et al., 2020). These scholars proposed that to effectively address misinformation, social media companies must consider whether they should post the factual correction before (prebunking) or after (debunking) users’ posts that contain damaging misinformation (Vraga et al., 2020). Additionally, the scholars noted that platforms must take into account the perceived credibility of the misinformation and also the credibility of the proposed correction (Vraga et al., 2020).
To test the effectiveness of user-based correction strategies on the social media platform, Instagram, the scholars conducted a study (Vraga et al., 2020). They began their study by assessing the placement of misinformation corrections in relation to the misinformation itself (Vraga et al., 2020). The scholars explained the inoculation framework as “(1) forewarning individuals about the potential of seeing a persuasive message and (2) providing them a weakened version of the persuasive message and the means to counterargue the message” (Vraga et al., 2020). In this technique, the platform displays a warning to users about the prevalence of misinformation and then exposes them to the misinformation (Vraga et al., 2020). In contrast, the correction framework uses corrective information after misinformation exposure (Vraga et al., 2020).

Vraga et al. (2020) also differentiated between fact-focused and logic-focused interventions (Vraga et al., 2020). They maintained that while fact-focused interventions counter misinformation by providing individuals with accurate information, logic-focused interventions involve explanations of the techniques used to mislead audiences (Vraga et al., 2020). While logic-focused corrections may use facts in countering misinformation, these facts explain the misleading techniques associated with the misinformation rather than refuting the misinformation itself (Vraga et al., 2020).

Vraga et al. (2020) found that logic-focused corrections were more effective in reducing users’ misperceptions than fact-focused corrections (Vraga et al., 2020). Both logic-focused and fact-focused corrections were equally effective in debunking misinformation, but only the logic-focused corrections reduced users’ misperceptions through “prebunking” (Vraga et al., 2020). One explanation for this finding is that logic-focused corrections reduce credibility ratings for the misinformation post through an indirect pathway that is not significant for fact-focused corrections (Vraga et al., 2020).

When assessing the credibility of misinformation corrections, Vraga et al. (2020) cited previous research from Nyhan and Reifler (2020), and Thornson (2016). These scholars’ research found that expert sources, including the news media, health organizations, government agencies, and fact-checking organizations, are beneficial in correcting online misinformation (Vraga et al., 2020).
In their study, Vraga et al. (2020) held source credibility constant and focused on message credibility (Vraga et al., 2020). In the study, participants rated fact-focused corrections higher in credibility than logic-focused corrections in all timing conditions, but especially when the participants saw the corrections prior to the misinformation (Vraga et al., 2020). The scholars explained this finding by stating that participants appreciated the fact-focused “prebunking” because it offered them valuable facts (Vraga et al., 2020). However, with logic-focused “prebunking,” participants saw this correction as being out of place when it appeared before the misinformation (Vraga et al., 2020). This finding showed that fact-focused corrections are more effective than logic-focused corrections when individuals and organizations hope to discredit misinformation (Vraga et al., 2020).

Overall, the findings of Vraga et al. (2020) slightly differ from those of Gesser-Edelsburg et al. (2018). Vraga et al. (2020) concluded that using logic-focused corrections that expose the misleading techniques misinformation employed were generally more effective than fact-focused corrections that counter misinformation with accurate information (Vraga et al., 2020). Unlike Vraga et al. (2020), Gesser-Edelsburg et al. (2018) concluded that participants rated theory-based information corrections, which provide information transparent information from health organizations to address the public’s emotional concerns, higher in reliability and satisfaction than information corrections that cite the misinformation as myths (Gesser-Edelsburg et al., 2018).

However, the disparity in results between these two studies arises out of the slight differences in each study’s conditions. First, the Vraga et al. (2020) study’s fact-focused corrections did not specifically address participants’ emotional concerns, (Vraga et al. 2020) while the Gesser-Edelsburg et al. (2018) study’s theory-based information corrections did address them (Gesser-Edelsburg et al., 2018). Vraga et al.’s (2020) fact-focused corrections merely countered misinformation with accurate information, but did not do so in a manner that addressed the emotions of participants (Vraga et al., 2020). Nevertheless, the Vraga et al. (2020) study did conclude that participants rated fact-focused as more credible than logic-focused corrections even though fact-focused corrections were less effective overall (Vraga et al., 2020).
Additionally, the Vraga et al. (2020) study’s logic-focused corrections included an explanation of the misinformation employed to mislead participants, (Vraga et al., 2020) while the Gesser-Edelsburg et al. (2018) study’s information corrections cited the misinformation as myths (Gesser-Edelsburg et al., 2018). By not repeating and reinforcing participants’ memories of the misinformation, the Vraga et al. (2020) study’s logic-focused corrections could be more effective than the Gesser-Edelsburg et al. (2018) study’s information corrections that allowed individuals to remember the misinformation (Gesser-Edelsburg et al., 2018).

**Considering Topic Familiarity and Information Credibility**

**Individual Familiarity.** Cowley and Janus (2004) studied whether advertising can alter consumers’ memories such that individuals remember a brand as being different than their actual, initial experiences with the brand (Cowley & Janus, 2004). In their study, the scholars exposed participants to misinformation about the type of juice they had consumed to determine whether advertising can alter a person’s memory of a central observation about a product (Cowley & Janus, 2004). In this study, Cowley and Janus (2004) assessed participants’ familiarity with a product. The scholars hypothesized that people who are quite familiar with a product would collect more information from experience and be less likely to incorporate consequent misinformation from an advertisement in their evaluations of the product than those who are less familiar with the product (Cowley & Janus, 2004).

When considering individuals’ susceptibilities to misinformation, Cowley and Janus (2004) cited the findings of Alba and Hutchinson (1987) that stated that familiarity allows consumers to accumulate, integrate, and judge the relevance of product information in a superior manner (Cowley & Janus, 2004). Cowley and Janus (2004) also addressed the findings of Hoch and Deighton (1989) that determined that familiarity results in an enhanced sophisticated memory schema and reduced suggestibility and susceptibility to memory distortion caused by exposure to post-experience misinformation (Cowley & Janus, 2004). Additionally, Cowley and Janus (2004) addressed the studies of Alba and Hutchinson (1987), and Melcher and Schooler (1996) that found that consumers familiar with products can better categorize these products to
create finer discriminations among similar products compared to consumers who are unfamiliar with the products (Cowley & Janus, 2004). Furthermore, Cowley and Janus (2004) asserted that individuals’ enhanced categorizations should reduce their susceptibilities to memory distortions (Cowley & Janus, 2004).

In their study, Cowley and Janus (2004) randomly assigned participants to two exposure conditions: one where participants viewed an advertisement with misinformation about the type of juice the participants had tasted, and one where participants did not view an advertisement (Cowley & Janus, 2004). While all participants sampled the juice, only the advertisement condition group viewed a misleading advertisement afterward (Cowley & Janus, 2004). Cowley and Janus (2004) then asked all of the participants to describe the juice, provide three words that best described their memories of the juice, and identify the juice they tasted out of a group of five juice samples (Cowley & Janus, 2004). The researchers found that participants in the advertisement condition group were no more likely than participants in the no-advertisement condition group to be susceptible to the misinformation about their central observations about the juice (Cowley & Janus, 2004).

However, Cowley and Janus (2004) found that within the group that the researchers exposed to the misleading advertisements, consumers who were less familiar with the juice were somewhat more likely to incorrectly remember the juice than those who were familiar consumers with the product (Cowley & Janus, 2004). Those participants who were familiar with the product and who were better at specifically categorizing products into subcategories were not only less susceptible to misinformation than other participants, but were also better at remembering accurate details about their experiences with the product because the misinformation had prompted their accurate recollections (Cowley & Janus, 2004). This finding supports the research of Schacter, Norman, and Koutstaal (1998), who determined that information that increases pattern separation can promote accurate reconstruction at the time of its retrieval (Cowley & Janus, 2004).
Source Credibility. Vraga and Bode (2017) studied the effectiveness of corrective responses to successfully reduce misperceptions during a misinformation outbreak by evaluating the impact of both the number of corrective responses and the credibility of their sources (Vraga & Bode, 2017). These scholars used the 2015 and 2016 Zika pandemic as the case study for their experimental design because of the widespread misinformation that accompanied the pandemic (Vraga & Bode, 2017). Regarding the Zika pandemic, Dredze et al. (2016) found that pseudo-scientific claims about the causes of the Zika virus and media attention on the virus on Twitter were correlated (Vraga & Bode, 2017). Additionally, Sharma et al. (2017) determined that posts that included misinformation about Zika were in the minority of posts on Facebook, yet these incorrect posts received the most public engagement (Vraga & Bode, 2017).

To support their hypothesis, Vraga and Bode (2017) used an experimental design that exposed users to misperceptions about the Zika virus’s causes on a simulated Twitter feed (Vraga & Bode, 2017). The scholars then introduced the study’s participants to a refutation of the misinformation that they manipulated in several ways according to various independent variables (Vraga & Bode, 2017). One independent variable the scholars manipulated was the correction source, which included various information sources (Vraga & Bode, 2017). The scholars changed the correction source to the Centers for Disease Control and Prevention (CDC), an unknown Twitter user, or both the CDC and an unknown Twitter user (Vraga & Bode, 2017). The scholars chose the CDC as one correction source because the American public commonly accepts the CDC as a reputable source of information compared to the average Twitter user as an information source (Vraga & Bode, 2017). The other independent variable they manipulated was the number of corrections to the misinformation (Vraga & Bode, 2017).

The Vraga and Bode (2017) study found no significant decrease in Zika misperceptions among users who viewed the user-only correction, compared to the control condition (Vraga & Bode, 2017). Meanwhile, the CDC-only correction significantly reduced participants’ misperceptions compared to the control condition and the user-only corrections (Vraga & Bode, 2017). In examining the impact of the number of corrections, the authors found that the number of corrections of misinformation reduced participants’ misperceptions compared to the control condition only when...
participants viewed a CDC correction after they had seen an unknown Twitter user’s correction to the misinformation (Vraga & Bode, 2017). In contrast, Vraga and Bode (2017) did not find a marginally significant difference between the control condition and the condition where they added a user correction following one from the CDC (Vraga & Bode, 2017).

Kim et al. (2020) offered an alternative approach to providing credible information corrections in response to misinformation on social media (Kim et al., 2020). Kim et al. (2020) proposed that organizations could use humor to remediate false claims on social media about the human papillomavirus (HPV) (Kim et al. 2020). Kim et al. (2020) asserted that humor might be an effective means to correct misinformation because of its ability to gain users’ attention at a more significant rate than other types of corrective information posts (Kim et al., 2020).

The scholars employed an experimental design in which they randomly assigned participants to one of two experimental conditions: humor and non-humor corrections (Kim et al., 2020). The participants viewed an inaccurate post on Twitter that claimed that the HPV vaccine caused auto-immune symptoms, such as paralysis (Kim et al., 2020). The scholars followed the post with a correction image and text, that varied depending on participants’ condition assignments (Kim et al., 2020). The scholars found that participants paid significantly more attention to the correction image in the humor condition than in the non-humor condition (Kim et al., 2020). However, they saw no significant differences in participants’ attention to the correction text in either condition (Kim et al., 2020). Kim et al. (2020) concluded that, overall, the non-humorous correction was more effective than the humorous correction in reducing misperceptions because participants rated the non-humorous corrections as more credible than the humorous corrections (Kim et al., 2020). The scholars found that when participants viewed the correction as being more credible, they reduced their HPV misperceptions “without an indirect pathway via attention” (Kim et al., 2020).
The findings of the Kim et al. (2020) study confirmed the conclusions from Vraga and Bode’s (2017) study. Vraga and Bode (2017) asserted that misinformation corrections are most effective when they come from a credible source, such as the CDC (Vraga & Bode, 2017). Kim et al.’s (2020) study expanded on this finding and concluded non-humorous corrections were most effective in reducing users’ misperceptions on social media because these types of corrections carried more credibility than humorous ones when correcting misinformation (Kim et al., 2020).

**Correction of COVID-19 Vaccination Misinformation**

When exploring strategies to increase COVID-19 vaccination uptake, French et al. (2020) found that the general population was experiencing a rise in vaccine hesitancy, which the scholars defined as “the delay in acceptance or refusal of vaccines despite the availability of vaccination services” (French et al., 2020). During the COVID-19 pandemic, the World Health Organization (WHO) recommended that “Each country should develop a strategy to increase acceptance and demand for vaccination” (French et al., 2020).

French et al. (2020) recommended that to combat this misinformation and encourage vaccination uptake, government and regional bodies must reduce the population’s concerns about the vaccines and encourage demand for the vaccines (French et al., 2020). The scholars believed that many individuals’ concerns that could reduce vaccine uptake stemmed from the online spread of misinformation and conspiracy theories about the COVID-19 pandemic, COVID-19 vaccines, and community vaccinations (French et al., 2020).

Additionally, Schoch-Spana et al. (2020) identified some of the many causes for the flood of misinformation during the COVID-19 pandemic. These causes included the widespread use of social media platforms as tools for information seeking, the uncertain nature of the novel disease, and the presence of misinformation campaigns that attempted to deflect blame and create false narratives around the global COVID-19 response (Schoch-Spana et al., 2020). Schoch-Spana et al. (2020) also claimed that speculation during the development of the COVID-19 vaccines contributed to the pool of misinformation (Schoch-Spana et al., 2020). The scholars noted that misinformation
regarding COVID-19 vaccines ranged from rumors about the vaccines’ safety to narratives that claimed that major organizations had created both the COVID-19 virus and its vaccines, with plans to use the vaccines for future financial gains (Schoch-Spana et al., 2020).

French et al. (2020) stated that to compete against the anti-vaccination movement and its use of misinformation, effective campaigns to support the COVID-19 vaccines should emphasize the potential dangers of the disease and highlight the benefits of the vaccines (French et al., 2020). The scholars believed that these tactics would motivate people by tapping into individuals’ fears of loss and their desires for good health (French et al., 2020). Additionally, they proposed that the campaigns should tailor vaccine promotions for different target populations, depending on each group’s specific motivators to receive vaccines (French et al. 2020).

Schoch-Spana et al. (2020) agreed with French et al. (2020) in asserting that to effectively engage with audiences with diverse beliefs and circumstances, and encourage uptake of the COVID-19 vaccines, authorities must engage with these audiences to understand their “concerns, values, attitudes, perceptions, and beliefs” (Schoch-Spana et al., 2020). According to the scholars, authorities can effectively communicate with their audiences by placing communities’ well-being at the center of communications and rejecting the politicized nature of the COVID-19 pandemic (Schoch-Spana et al., 2020).

Additionally, Schoch-Spana (2020) explained that the communications should come from a broad network of trusted spokespersons to ensure that the messages spread throughout communities (Schoch-Spana et al., 2020). Schoch-Spana et al. (2020) noted that although trusted spokespersons would be necessary to convey the messages, the messages’ virality would be significant because messages shared among members of a community may carry more substantial weight than those exclusively from officials or spokespersons (Schoch-Spana et al., 2020).

Furthermore, French et al. (2020) believed that “fact-checking and myth-busting” misinformation may not be effective because these tactics often pass along the misinformation, and further instill it in individuals’ minds (French et al., 2020). The scholars explained that engaging directly with misinformation further spreads the erroneous information rather than eliminating it because of individuals’ susceptibilities to
the confirmation bias (French et al., 2020). Therefore, when sources address
misinformation in an attempt to disprove it, individuals might be receptive only to the
misinformation, and not believe the opposing claim (French et al., 2020). The scholars
suggested that, instead, authorities should encourage vaccination uptake with a
“combination of positive messaging that emphasizes the protective (individual, family,
and community) benefit of the vaccine and the loss associated with not being vaccinated
(death, poor health, loss of freedom and social solidarity, inability to travel, etc.)” (French
et al., 2020). French et al. (2020) supported this strategy with the findings of Tversky and
Kahneman (1986) that stated that individuals are averse to choices that could result in a
perceived loss (French et al., 2020).

French et al. (2020) also stressed that public health authorities should not permit
anti-vaccination online users to spread misinformation continuously (French et al., 2020). The
scholars stated that authorities should work with media sectors to proactively reduce
and remove false content and misleading information (French et al., 2020). They
recommended that health authorities form agreements with media companies to identify,
flag, and remove false and misleading anti-vaccination information (French et al., 2020).
They believed that through this approach, health authorities would have access to the
misinformation that traditional media and social media sectors had detected (French et
al., 2020).

Summary

This literature review evaluates scholars’ varied views regarding the ethicality of
censorship; the use of third-party regulators to eliminate misinformation; the means of
classifying misinformation; the explanations for the spread of misinformation; the
impacts of misinformation on individuals and social media platforms; and the methods of
correcting misinformation, both generally and during the COVID-19 pandemic.

When evaluating the use of censorship and its impact on individuals’ freedom of
speech, scholars differ in their opinions on the degree to which the media should censor
consumers’ content. The scholars agree that freedom of speech is necessary to facilitate
social exchanges of ideas to counter false ideas with correct information (Harer &
Harrell, 2002). However, the scholars disagree about the role the media should assume in
censoring content. Siebert et al. (1984) expanded on Mill’s philosophy that censorship limits the necessary flow of information among individuals by claiming that the press must be responsible to society by not limiting the public’s information (Harer & Harrell, 2002). However, Emerson, a Constitutional scholar, stated that the press should use censorship when the speech encourages actions that destroy public order (Harer & Harrell, 2002). Niesien (2020) disagreed with Emerson’s perspective, stating that categorizing information is challenging and individuals often interpret information subjectively (Niesien, 2020).

Moreover, when addressing what constitutes “misinformation,” scholars differ significantly in their classifications and definitions. Schoch-Spana et al. (2020) defined misinformation as “information that differs from expert consensus at the time it is shared” (Schoch-Spana et al., 2020). Meanwhile, Niemiec (2020), and Chou et al. (2018) agreed that individuals can identify health misinformation by its lack of scientific evidence (Chou et al., 2018). However, Niemiec (2020) further defined misinformation by distinguishing between health and political misinformation (Niemiec, 2020). In contrast, Wang et al. (2019) differentiated misinformation from disinformation, stating that individuals do not create misinformation to cause harm, while they do have this malintent with disinformation (Wang et al., 2019). Furthermore, Lazer et al. defined the commonly used term, “fake news,” as “fabricated information that mimics news media content” (Wang et al., 2019). However, Wang et al. claimed that Lazar et al.’s (2018) definition did not encompass the entirety of misinformation that can also include satire and information created to achieve an alternative end, such as a political goal (Wang et al., 2019).

After evaluating what scholars believe constitutes misinformation, the literature review compares the social media industry’s censorship policies to those of other industries. The scholars noted that while other industries, such as the prescription drug industry, have programs to limit the spread of misinformation and hold companies accountable if they contribute to its spread, the social media industry does not have similar programs in place. Despite the industry’s lack of oversight, Niemiec (2020) disagreed with the viewpoint that the social media industry must increase its censorship of misinformation. He proposed that social media platforms already have overbearing
control over the content to which viewers have access (Niemiec, 2020). In fact, Niemiec (2020) asserted that social media platforms can shape users’ points of view by exposing them to targeted content; thus, increased censorship would further harm the industry rather than solve the problem of widespread misinformation (Niemiec, 2020).

The scholars also addressed how health-related misinformation spreads on social media and examined the root causes behind its distribution. For example, Wang et al. (2019) cited the “basic law of rumor” as contributing to the spread of misinformation and explained that the rate at which information spreads depends on individual perceptions regarding the importance of the information, as well as the ambiguity of evidence relevant to the information (Wang et al., 2019). Chou et al. (2018) proposed a different cause for the spread of misinformation. They claimed that information silos and echo chamber effects amplify the spread of misinformation (Chou et al., 2018). The scholars contended that individuals rely on these biases and consequently use targeted posts that align with their biases to confirm their previous beliefs (Chou et al., 2018). Chou et al. (2018) added that individuals’ lack of trust in institutions, such as the medical community, increases the spread of health-related misinformation (Chou et al., 2018).

Niemiec (2020) agreed with Chou et al. (2018) on their notion that human psychology influences the spread of misinformation on social media. Niemiec (2020) expanded on this idea by stating that social media sites depend on humans’ needs for social approval, reciprocity, and novelty seeking for the success of the platforms’ targeted posts and information-screening activities (Niemiec, 2020).

Moreover, most scholars agreed that misinformation could create harmful impacts for both individuals and companies. Chou et al. (2020) addressed the distrust that individuals have in the medical community as a result of the increased prevalence of health-related misinformation and a lack of that community’s unified response to crises (Chou et al., 2020). These scholars stated that distrust in science and the healthcare community could lead individuals to ignore health information; avoid health care; or make uninformed, detrimental health decisions (Chou et al., 2020). Furthermore, Grygiel and Brown (2019) addressed the impacts of misinformation on social media platforms’ brand images and profitability and added that social media firms need content moderation to maintain their user bases and marketability to advertisers (Grygiel & Brown, 2019).
After establishing scholars’ consensus about the necessity of removing misinformation from social media websites, the literature review evaluates the scholars’ varied perspectives about the appropriate means of correcting misinformation. Many scholars agreed that education is necessary to inform users about their potential exposure to misinformation. However, scholars differed in what they believed was the proper type of education for users. Chou et al. (2018) stated that users should be educated about the medical community to increase their trust in the field and encourage them to be receptive to releases of information from those sources (Chou et al., 2018). Niesien (2020), however, cautioned that users should also be aware of the lack of reliability of some scientific organizations, explaining that such information may not be free from errors, and social media platforms should not propose these organizations as definitive sources for accurate information (Niesien, 2020). Additionally, Chou et al. (2018) encouraged social media platforms to determine which populations are the most vulnerable, assess the reach of misinformation into these communities, and determine the potential health consequences of this misinformation (Chou et al., 2018).

Next, Chou et al. (2020) added that correction measures must overcome the human confirmation bias, stating that value confirmation might be a more effective correction method than corrective information (Chou et al., 2020). However, Bode and Vraga (2015) challenged the strength of the human confirmation bias by studying the use of suggested posts that contained corrective information on Facebook (Bode & Vraga, 2015). These scholars concluded that exposure to corrections from related stories to the misinformation on social media could decrease users’ misperceptions (Bode & Vraga, 2015). This finding demonstrated that corrective information could alter users’ pre-existing attitudes (Bode & Vraga, 2015).

The literature review then analyzes various types of information corrections. First, Gesser-Edelsburg et al. (2018) evaluated the effectiveness of two forms of misinformation corrections on health-related misinformation: information corrections and theory-based information corrections (Gesser-Edelsburg et al., 2018). The scholars defined information corrections as posts that cite misinformation as myths and theory-based information corrections as posts that provide transparent information from health organizations that address the public’s emotional concerns (Gesser-Edelsburg et al.,
Gesser-Edelsburg et al. (2018) found that participants rated theory-based information corrections higher in reliability and satisfaction than information corrections (Gesser-Edelsburg et al., 2018). These findings showed the importance of correcting misinformation transparently and overcoming the human confirmation bias with emotional appeals in theory-based information correction (Gesser-Edelsburg et al., 2018). Finally, Gesser-Edelsburg et al. (2018) concluded that information corrections are ineffective because they reinforce individuals’ memories of misinformation, even when evidence exists that establishes it as false (Gesser-Edelsburg et al., 2018).

Similarly, Vraga et al. (2020) explored fact-focused and logic-focused interventions to correct misinformation (Vraga et al., 2020). These scholars explained that fact-focused interventions counter misinformation with accurate information, while logic-focused interventions explain the misleading techniques associated with persuasive misinformation (Vraga et al., 2020). Vraga et al. (2020) found that logic-focused corrections and fact-focused corrections were equally effective when used before misinformation exposure (Vraga et al., 2020). However, only logic-focused corrections reduced users’ misperceptions when they appeared following misinformation exposure (Vraga et al., 2020). Despite their responses to logic-focused corrections, participants rated fact-focused corrections as more credible than logic-focused corrections, especially when they saw them prior to the misinformation (Vraga et al., 2020).

The findings of Vraga et al. (2020) slightly differed from those of Gesser-Edelsburg et al. (2018) because according to Vraga et al. (2020), corrections that explain the deception techniques of the misinformation are more effective than corrections that provide accurate information that contradicts the misinformation (Vraga et al., 2020). Gesser-Edelsburg et al. (2018) determined that corrections that provide transparent information from health organizations are more effective than corrections that label the misinformation as myths (Gesser-Edelsburg et al., 2018).
However, the discrepancies in the conditions of the two studies may explain these differences. First, the Vraga et al. (2020) study’s fact-focused corrections did not address the participants’ emotional concerns, but solely provided contradicting accurate information to counter the misinformation (Vraga et al., 2020). Additionally, the Gesser-Edelsburg et al. (2018) study’s information corrections did not explain the techniques the originators of the misinformation used to circulate it (Gesser-Edelsburg et al., 2018). Instead, the scholars merely labeled the misinformation as myths (Gesser-Edelsburg et al., 2018).

Next, the literature review evaluates the influence of familiarity and information credibility on the effectiveness of misinformation corrections. Cowley and Janus (2004) studied how individuals’ familiarity with a subject affected their susceptibility to misinformation. The scholars found that participants who were less familiar with a subject and were exposed to post-experience misinformation advertisements were more likely to have memory distortions than those participants familiar with a subject (Cowley & Janus, 2004). Cowley and Janus (2004) concluded that individuals with subject familiarity have more sophisticated memory schemas to categorize information about a subject than individuals with less familiarity (Cowley & Janus, 2004). The scholars contented that this improved categorization ability allows consumers with some subject familiarity to make finer discriminations about similar products and be less likely to incorporate misinformation into their evaluations than consumers with less familiarity (Cowley & Janus, 2004).

When evaluating the influence of source credibility on misinformation corrections, Vraga and Bode (2017) exposed users to misinformation about the causes of the Zika virus (Vraga & Bode, 2017). They then introduced misinformation corrections from various sources, including an unknown Twitter user, the CDC, or both (Vraga & Bode, 2017).
The scholars concluded that CDC corrections significantly reduced users’ misperceptions in two conditions: the CDC-only, and the combined CDC and unknown Twitter user. (Vraga & Bode, 2017). These findings notwithstanding, the scholars did not observe a decrease in misperceptions when participants viewed an unknown Twitter user’s corrections after seeing misinformation (Vraga & Bode, 2017). This conclusion highlighted the necessity for misinformation corrections to come from reliable sources if they are to be effective in changing users’ misperceptions (Vraga & Bode, 2017).

Kim et al. (2020) expanded on Vraga and Bode’s (2017) conclusion that the reliability of the corrective source is important in correcting users’ misinformation perceptions. Kim et al. (2020) determined that participants rated non-humorous corrections as being more credible and more effective at reducing users’ misperceptions than humorous corrections (Kim et al., 2020). However, the scholars differed in their approaches to studying the factors that contribute to a corrective claim’s effectiveness. While Kim et al. (2020) studied the importance of using an emotional appeal to correct misinformation claims, Vraga and Bode (2017) evaluated the importance of the credibility of the source that was distributing the corrective information.

Finally, the literature review assesses misinformation correction techniques specific to the COVID-19 pandemic and the release of COVID-19 vaccines. French et al. (2020) proposed that governments and health authorities can increase COVID-19 vaccination uptake by reducing the public’s concerns and correcting misperceptions about the vaccines (French et al., 2020). These scholars asserted that COVID-19 vaccination campaigns should focus on the target populations’ specific motivators for receiving the vaccines and emphasize the potential losses associated with not receiving the vaccines (French et al., 2020).
Schoch-Spana et al. (2020) agreed with French et al. (2020) in that authorities must engage with target audiences to understand their concerns, values, attitudes, perceptions, and beliefs (Schoch-Spana et al., 2020). Furthermore, Schoch-Spana et al. (2020) added that authorities should place communities’ well-being at the forefront of their vaccination communications to reduce the influence of politics on individuals’ vaccination decisions (Schoch-Spana et al., 2020). These scholars also noted that a broad network of trusted sources should distribute pro-vaccination communications and encourage local community members to share the information with their neighbors (Schoch-Spana et al., 2020).

Findings for Correcting COVID-19 Vaccines and Vaccination Misperceptions

Changing Consumers’ Beliefs About the Importance Of Vaccinations

When attempting to reduce public misperceptions about the COVID-19 vaccines that resulted from widespread misinformation, public health and governmental authorities must change users’ perceptions regarding the vaccines and their beliefs about the importance of the vaccinations.

To raise users’ levels of attention to misinformation corrections related to the COVID-19 vaccines, authorities must first increase the topic’s importance to users. As Wang et al. (2019) noted, the “basic law of rumor” states that the amount of rumor spread depends on both individuals’ perceptions about the importance of the subject and the ambiguity of relevant evidence about the subject (Wang et al., 2019). In accordance with this principle, public health officials must increase the importance of the vaccines in users’ minds to increase their receptivity to new, accurate information corrections regarding the COVID-19 vaccines and vaccinations.
Wang et al. (2019) further explained that a message’s ambiguity increases when individuals receive conflicting information from equally-credible sources (Wang et al., 2019). In the case of COVID-19 vaccines, public health authorities can reduce ambiguity about the vaccines for users by releasing accurate information from highly-credible sources that users will perceive as being more reliable than previous, ambiguous user-generated information on social media platforms. By increasing the perceived importance of the COVID-19 vaccines and vaccinations to users and decreasing the ambiguity of credible information, authorities can effectively disseminate information corrections that will be effective in reducing previous misperceptions.

French et al. (2020) stated that authorities should encourage vaccination uptake through messages that emphasize both the benefits of the vaccines and the potential costs of not receiving the vaccines (French et al., 2020). French et al. (2020) cited the findings of Tversky and Kahneman (1986), who stated that individuals are averse to choices that could result in perceived loss (French et al., 2020). In fact, COVID-19 has many potential detrimental effects on individuals, including poor health, loss of freedom and social solidarity, inability to travel, or death (French et al., 2020). The scholars believed that by emphasizing the adverse outcomes of failing to receive the COVID-19 vaccine through loss-framed messages, health officials can effectively incentivize vaccination uptake by the public (French et al. 2020). In accordance with these findings, authorities should consider promoting the importance of the COVID-19 vaccines and vaccinations by using more loss-framed messages and fewer gain-framed messages.

Furthermore, public health authorities must change users’ beliefs about the COVID-19 vaccination to effectively correct the public’s misperceptions. According to Chou et al. (2020), changing individuals’ beliefs can be challenging because the human confirmation bias encourages individuals to accept information that confirms their current beliefs and reject information that opposes those beliefs (Chou et al., 2020). However, Bode and Vraga (2015) determined that while individuals give low evaluations to information that opposes their current misperceptions, these low evaluations do not prevent new information from changing individuals’ attitudes (Bode & Vraga, 2015). Therefore, presenting opposing information can alter individuals’ beliefs, even in the presence of the human confirmation bias.
Moreover, to further overcome the human confirmation bias, Chou et al. (2020) suggested using value affirmation interventions to correct misinformation (Chou et al., 2020). Value affirmation interventions are similar to the theory-based information corrections proposed by Gesser-Edelsburg et al. (2018). Gesser-Edelsburg et al. (2018) found that individuals viewed theory-based information corrections, which include transparent information from health organizations that address the public’s emotional concerns, as more reliable and satisfactory than information corrections that merely cite the misinformation as myths (Gesser-Edelsburg et al., 2018).

Cowley and Janus (2004) elaborated on these findings by asserting that authorities must also increase users’ familiarity with the topic to reduce their susceptibility to memory distortions from misinformation (Cowley & Janus, 2004). Cowley and Janus (2004) found that familiarity with a subject results in an augmented, complex memory schema and categorization process among individuals (Cowley & Janus, 2004). The scholars determined that increased categorization allows individuals to remember accurate details and avoid confusion from misinformation (Cowley & Janus, 2004). In the case of the COVID-19 pandemic, by improving users’ knowledge about the COVID-19 vaccines and vaccinations through accurate data about their safety and efficacy, health officials can reduce users’ vulnerabilities to social media misinformation.

**Establishing Source Credibility**

Vraga and Bode (2017) asserted that corrective posts can reduce users’ misperceptions if the posts come from a credible source, such as the CDC (Vraga & Bode, 2017). These scholars found that users’ misperceptions did not decrease significantly when a correction came from a non-authoritative user (Vraga & Bode, 2017). This experimental study revealed the increased effectiveness of using a credible source for information corrections rather than an unknown source, such as a member of the general population (Vraga & Bode, 2017). Similarly, Kim et al. (2010) found that non-humorous corrections were more effective in correcting false claims about HPV than humorous corrections because participants rated the non-humorous corrections high in credibility (Kim et al., 2010). While the participants in the Kim et al. (2010) study paid more attention to the correction image in the humorous condition than one in the non-
humorous condition, they did not evaluate the humorous correction images as being credible (Kim et al., 2020). Both Vraga and Bode (2017), and Kim et al. (2010) established misinformation corrections as being increasingly effective in reducing misperceptions when users view the corrections as being credible.

Schoch-Spana et al. (2020) agreed with the perspectives of Vraga and Bode (2017), and Kim et al. (2020) who stated that corrective communications should be credible (Schoch-Spana et al., 2020). However, Schoch-Spana (2020) also added that the virality of the corrections is another important factor in reducing users’ misperceptions (Schoch-Spana et al., 2020). These scholars asserted that messages shared throughout individual communities may be more influential than those that come only from officials or spokespersons (Schoch-Spana et al., 2020). Schoch-Spana (2020) believed that, therefore, the correction communications should originate from trusted spokespersons, but for maximum efficacy, users within their respective communities should share and repost the messages (Schoch-Spana et al., 2020). Health officials can apply these findings directly to COVID-19 vaccination communications, especially as campaign managers select their sources.

**Recommendations for Future Research**

One area of research that needs further development within the subject of misinformation, and specifically health-related misinformation, concerns the prevalence of misinformation and the trends in how it spreads. While misinformation can be challenging to accurately track because the creators of misleading content are motivated to mask misinformation as accurate information, scholars should explore the trends that emerge during misinformation outbreaks. These explorations could include the demographic, geographic, psychological, and behavioral characteristics that make consumers increasingly susceptible to developing misperceptions following misinformation exposure.
Additionally, future researchers should examine the content creators of disinformation to identify their political, social, economic, or other incentives. The insights researchers receive from these analyses will allow policymakers and social media platforms to understand how to effectively combat disinformation outbreaks and dismantle any underlying drivers for the creators of this harmful content.

Scholars should also conduct further research on the impacts misinformation has on consumers, and the resulting consequences for companies’ brand images and customer loyalties. Future researchers should investigate the impact that content filtering, or lack thereof, has on consumers’ perceptions of a social media firm. This research would include analyses of companies’ corporate social responsibility (CSR) initiatives and these initiatives’ effects on consumer segment revenues, customer lifetime value (CLV), and brand loyalty. These findings will help industry analysts to understand how companies’ approaches to addressing misinformation can produce varied outcomes. The results of this research will be especially important for companies to determine whether to use censorship. Censorship will deflect immediate revenues from increased user engagement and advertising in an effort to protect consumers. The findings of research into content filtering will help companies to decide whether protecting consumers through censorship will be ultimately profitable by increasing brand loyalty and sparking new customer acquisition.
REFERENCES


Social Media Censorship: The Impact Filtering Information Has on the Well-Being of Users and Online Businesses

Case Overview

This instructor’s manual will exemplify the effects of social media censorship on businesses and individual communities. The case study is framed from the viewpoint of Mike Schroepfer, Chief Technology Officer (CTO) of Facebook.

The case study will explore Facebook’s history as a pioneer social media network. Mark Zuckerberg, a college student at Harvard University at the time, launched Facebook in 2004 (Boyd, 2019). The social media website was instantly popular among college students hoping to connect online. By the end of 2007, Facebook memberships had grown to over 50 million users (Boyd, 2019).

With its newly-pronounced influence across the globe, Facebook had to make challenging decisions to define its power boundaries. Specifically, Facebook began instating its influence as a political force with the 2008 Presidential election. At this time, Facebook encouraged users to vote in the historic election, asserting its dominance as a prominent news outlet (Dutta & Fraser, 2008). This power expanded to the platform’s use of targeted advertisements. In 2011, Facebook launched a campaign that distributed targeted content at specific demographics to oppose a Florida ballot proposition (Facebook, 2011).

As Facebook’s influence over consumers’ lives continued to increase, it faced additional struggles in balancing its users’ safety and protection with remaining as a platform of free speech for users. Beginning in 2015, the company received rising criticisms from users and officials over the prevalence of misinformation on the company’s website (Boyd, 2019).
To resolve these concerns, Facebook installed new features that allowed users to flag questionable stories as “false news stories” and began filtering and removing flagged content from the platform (Boyd, 2019). Despite these efforts, misinformation continued to spread on the website, often receiving more attention than factually-accurate content posted on the platform. In 2017, Facebook further increased its preventative messages against misinformation by enacting a three-step approach (Facebook, n.d.). These initiatives included disrupting economic incentives for posting misleading information, helping people make informed decisions through user education, and building new technology products to filter false information and prioritize accurate posts (Facebook, n.d.)

Nevertheless, Facebook’s efforts remained flawed. The company still must determine the appropriate measures to balance reducing misinformation with not eliminating opinion pieces from the platform. Mike Schroepfer, Facebook’s CTO who leads the company’s artificial intelligence (AI) efforts, must now determine the most effective means to safeguard users on the platform regarding their exposures to misinformation and their rights to freedom of speech.

This case study’s research includes varied news sources, scholarly articles, and interviews with Facebook’s company leadership. The case study requires minimal background knowledge about the company but does need readers to have a basic understanding of social media operations in the early twenty-first century.

Intended Audience, Recommended Courses, and Placement

The intended audience for this case is undergraduate students who are studying marketing or management. These students will hopefully have an academic background in social and online marketing, as this case reviews the marketing techniques of a social media company, Facebook. Additionally, these undergraduate students should have general knowledge about the COVID-19 pandemic and its global impact from 2020 to 2021.

Additionally, the case is relatively short, making it appropriate for undergraduate students. Students should study the case in their fourth year of college. Students should
have solid foundational knowledge of marketing concepts and be entering discussion-led courses that will be necessary to facilitate discussion of the case.

At Tulane University, the course is well suited for Social and Online Marketing, Consumer Behavior, and Leadership. These courses are appropriate for the case because they focus on marketing in online settings, consumer responses to marketing efforts, and effective leadership through various challenges, respectively.

Furthermore, instructors should use the case at the beginning of the semester because it introduces the complex nature of online marketing and information censoring. Instructors can use this case study as a solid introduction for a course because it exemplifies the foundational concepts necessary to understand online marketing.

**Learning Objectives**

Students should be familiar with online information censoring and misinformation corrections on social media after completing and analyzing the case. Additionally, students should understand the importance of limiting misinformation spreads on social media to protect both individual users and social media platforms. They should also acknowledge the intense dilemma that social media companies face in filtering misinformation on social media and how their actions influence their brand identities.

**Discussion Questions**

The following questions address the critical learning objects of the case. Each learning objective has related questions to encourage further discussion about the material.
Learning Objective 1: Review Facebook’s current and future objectives for removing misinformation from its platform and assess how its current information filtering policies align with these goals.

“What is Facebook currently doing to proactively and retroactively reduce online misinformation spreads on its platform?”

Regular Student Answer:
Facebook’s current misinformation policies include disrupting economic incentives, building new products, and helping people make informed decisions. Additionally, Facebook’s CTO, Mike Schroepfer, leads a team at Facebook that works on developing new artificial intelligence (AI) technologies that will effectively detect and filter posts that contain misinformation.

Outstanding Student Answer:
Facebook has implemented various initiatives to prevent the creation and distribution of misinformation on its platform. To do so, in 2017, Facebook directed its resources towards implementing new company-wide initiatives that remain active currently.

The first area of focus for Facebook is to disrupt users’ economic incentives for creating misleading or inaccurate content. Facebook has made progress towards this objective by using advanced algorithms, tightening its advertising requirements, and collaborating with third-party fact-checking organizations.

For the second component of its plan, Facebook has created educational programs to help inform the public about how misinformation spreads, ways to detect false news, and what to do when reading inaccurate information. These programs include educating social media users through the Facebook Journalism Project and working with other industry leaders to inform communities through the News Integrity Initiative.

Facebook has recently shifted its attention towards developing new technologies that will assist in detecting and filtering inaccurate and misleading content from the
platform. Specifically, the company has focused on advancing its use of artificial intelligence (AI) technologies to slow the spread of misinformation.

However, despite these initiatives, Facebook struggles to articulate a clear stance on misinformation spreads and post-filtering and remains ambiguous in its approach to controversial topics, such as the COVID-19 vaccines.

“How did Facebook’s current policies for removing misinformation change during the COVID-19 pandemic and subsequent misinformation outbreaks?”

Regular Student Answer:

Facebook’s censorship policies changed following the COVID-19 pandemic outbreak. In an attempt to promote accurate COVID-19 related information, Facebook decreased its tolerance for misinformation on the platform and increased its efforts to flag and remove inaccurate information. Additionally, the company collaborated with third-party authoritative sources, such as the Center for Disease Control (CDC). Working with these authoritative sources, Facebook provided its users with reliable and credible health-related information to correct online misinformation and disinformation.

Outstanding Student Answer:

After the initial outbreak of the COVID-19 virus in 2020, Facebook experienced a surge of misinformation posts from users who were spreading false claims about the virus’s origins, potential cures, and potential harm. This misinformation outbreak forced Facebook to deal with this issue efficiently and promptly. In fact, according to the company, its priority was to provide “accurate and authoritative information” to users (Lewis, 2020). To meet this goal, Facebook worked closely with the Center for Disease Control and Prevention (CDC) to provide information corrections to misinformation posted on the platform.

Additionally, the company removed millions of misleading and inaccurate posts from the platform during the COVID-19 pandemic. However, during this process, the company struggled to balance its position as a free speech vehicle and a provider of accurate information for users.
Learning Objective 2: Understand the necessity of removing misinformation and assess the problems with filtering users’ content.

“How does misinformation negatively impact Facebook’s brand image, and how does filtering false or harmful information help Facebook’s brand equity?”

Regular Student Answer:

Misinformation negatively impacts Facebook’s brand image because it reduces the company’s information reliability. Online users will be less willing to frequent the Facebook platform and read news from the website if they feel that the information will be inaccurate or misleading.

Additionally, filtering harmful information on Facebook will help the platform to establish substantial brand equity by allowing users to establish trust in the company and feel safe using Facebook as a news source over competitors who may lack any form of censorship.

Outstanding Student Answer:

The presence of misinformation on Facebook is concerning for a variety of reasons, most notably because it could threaten users’ safety. Nevertheless, misinformation on the platform is also detrimental to Facebook’s brand equity, customer equity, and, ultimately, customer lifetime value (CLV). By having misinformation on its website, Facebook would create an environment that is no longer suitable for reliable news sources to advertise their information. Both advertisers and major news networks will stop purchasing space on the platform when the amount of misleading and inaccurate information exceeds that of factually accurate information. Given that Facebook generates most of its revenues from advertisements and users’ engagement with these advertisements, the company will suffer financially and lose its brand equity over other social media platforms if users perceive information on the platform to be unreliable.

Facebook can improve its brand equity by filtering harmful or false content on its platform. By doing so, the company will establish trust among its users and can establish itself as a reliable online news source.
“What risks does Facebook face by limiting users’ rights to complete freedom of speech on the platform?”

Regular Student Answer:

By limiting users’ rights to complete freedom of speech, Facebook risks legal challenges because individuals and companies may sue Facebook by claiming that the platform deprived them of free speech. Additionally, the company may lose users if these users feel they cannot freely voice their opinions on the platform.

Outstanding Student Answer:

As a private corporation, Facebook may face both legal and ethical challenges when filtering information. The company must ensure that it is not limiting users’ rights to freedom of speech when filtering content. Specifically, Facebook must carefully differentiate between opinion pieces and content that users post as purported factual information. This differentiation can be challenging for Facebook because individuals often interpret information subjectively.

Additionally, Facebook could lose some of its customer equity because users might stop frequenting the platform if it does not serve as an unregulated discussion board for opinions.
Learning Objective 3: Analyze Facebook’s dilemma during the COVID-19 pandemic and address how the company could handle misinformation during this time.

“How did the COVID-19 pandemic influence the spread of online misinformation throughout the social media industry and how did the pandemic impact the responses from various online companies?”

Regular Student Answer:

The COVID-19 pandemic increased the spread of misinformation on social media platforms because individuals felt uncertain and anxious and were willing to spread any information they heard that might bring them perceived clarity about the virus outbreak. In response, social media platforms had to be hyper-focused on reducing misinformation on their platforms.

Outstanding Student Answer:

The COVID-19 pandemic created an unprecedented amount of panic and uncertainty around the globe. The pandemic shifted the way individuals lived their lives and interacted with their peers. Consequently, the spread of misinformation on social media vastly increased concurrent with users’ use of social media platforms as means of gathering information and communicating with others.

In response, social media companies faced the challenging situation of having to reduce misinformation and provide accurate information to users on a constantly evolving health topic. This situation was especially complicated because scientists were frequently discovering new information about COVID-19 and replacing previous theories. Additionally, these platforms had to advance their technologies to keep up with the increased demand on social media and the limited supply of internal human content moderators. Facebook also made these decisions during a politically tumultuous time that left the country divided on many topics, including the COVID-19 vaccine uptake and distribution methods.
“If you were Schroepfer, how would you handle removing misinformation on Facebook, and specifically, balance the risk of unintentionally removing accurate information against the risk of failing to identify inaccurate information? Would you need any additional information to make an informed decision?”

*Regular Student Answer:*

Schroepfer should continue to reduce misinformation on Facebook’s platform through proactive and reactive filtering technologies and algorithms. Schroepfer should do so because if he does not, the company’s potential costs associated with having misinformation spread on Facebook will outweigh any benefits the company could gain by allowing users to post freely and unregulated on the platform.

*Outstanding Student Answer:*

Schroepfer has the unique problem of having to handle misinformation during a global pandemic. When looking at the complexity and importance of online information distribution during the COVID-19 pandemic, Schroepfer must determine how to protect Facebook users’ safety, even if doing so will come at a cost to the company’s profits.

Schroepfer should establish policies that have minimal tolerance for misleading or inaccurate information. While these strict policies might cause the platform to erroneously delete some users’ opinion pieces, especially when users phrase them as information posts, this method will ensure that users do not encounter harmful misinformation or disinformation that will perpetuate rumors during the already-troubling times. Additionally, Schroepfer should increase Facebook’s budget for upgraded technology for AI development and for additional human content moderators.

However, to make this decision, Schroepfer should review a detailed analysis of the company’s financials. This report would include the company’s daily user activity numbers during the pandemic and its weekly revenues following its new censoring policies and filtering initiatives.
Teaching Strategies

Case Preparation

Prior to students reading the case, the instructor should make sure that students understand the fundamentals of the marketing process. Students should understand how companies analyze the marketplace and their customer needs and wants; design customer value-driven marketing strategies; construct integrated marketing programs that deliver superior value; engage customers, build profitable relationships, and create customer delight; and capture value from their customers to create profits and ensure customer equity. This knowledge will be necessary for students to understand social media companies’ underlying motivations and their respective target market segments.

Before the students analyze and evaluate the case, the instructor should lead a discussion about the various ways in which social media companies can be profitable and students should write one paragraph about Facebook’s business model in terms of generating revenue.

Teaching Narrative

When helping students to understand the case study, instructors should compare and contrast Facebook’s motivations in operating the platform and users’ motivations for using the platform. Instructors can use a Venn diagram to show how these motivations differ and where they overlap. Additionally, instructors can create timelines with their students to follow the progression of Facebook’s censorship policies in relation to various misinformation outbreaks.

Instructors can pose the following two introductory questions to students prior to presenting the case:

- How many of you have seen a false news story on Facebook or another social media platform?
- Of those of you who raised your hands, how many of you have seen a correction to or warning about this misinformation on that social media platform?

These questions will prompt students to begin thinking about the widespread nature of misinformation on social media and the various approaches these companies have taken to limit its spread.
If instructors plan to use storyboards for students to follow when reading the case, they can use the following design. Instructors should break down the problem into three components and encourage students to share their input while reading the case study.

**FACEBOOK**

**THE DECISION:**
**SHOULD FACEBOOK FILTER THE MISINFORMATION ON ITS PLATFORM?**

**THE COMPANY**

What do we know about Facebook?
- Mark Zuckerberg founded Facebook in 2004 as one of the first social media platforms
- The company faced criticism about its post filtering and its contribution to online misinformation outbreaks
- Facebook hired Michael Schroepfer as its CTO in 2013 to help develop the company's technology initiatives

**THE CTO**

What do we know about Michael Schroepfer?
- Schroepfer co-founded a software company, CenterRun, that Sun Microsystems acquired and he was Mozilla's Vice President of Engineering
- He acknowledged the problems with Facebook's censorship policies and pledged to improve them
- Schroepfer headed the company's AI initiatives and Applied Machine Learning teams

**THE MISINFORMATION**

What do we know about misinformation on Facebook?
- Facebook had a long-standing history with misinformation outbreaks on its platform
- The company has dealt with political, health-related, and social misinformation on its platform and implemented various initiatives to prevent their spread
- In 2020, Facebook faced new challenges with health-related misinformation during the outbreak of the COVID-19 pandemic
Literature Review, Theory, and Recommended Readings

The case study uses Facebook as an example to illustrate the challenging dilemma social media platforms face when they attempt to filter users’ posts to reduce misinformation while still allowing users to post their opinions freely on the platform. The literature review expands on the case study by exploring the spread of online misinformation and studying the effects of using censorship to limit its reach. Students should use the literature review to understand scholars’ definitions of misinformation, their explanations for its spread, and their solutions for reducing online users’ exposure to it. The literature review focuses on health-related misinformation and the COVID-19 pandemic, specifically.

Schoch-Spana et al. (2020) examined the spread of online misinformation during the development and uptake of the COVID-19 vaccines. These scholars suggested that the widespread online misinformation spread was the result of several factors. The researchers pointed to individuals using social media platforms as information-seeking tools, uncertainty amongst communities regarding the nature of the disease, and online users’ implementation of misinformation campaigns that intended to deflect blame and create false narratives about the pandemic (Shoch-Spana et al., 2020). To combat this misinformation and encourage online users to receive the COVID-19 vaccines, French et al. (2020) recommended that governments reduce the population’s concerns about the vaccines and increase demand for the vaccines by emphasizing the potential dangers of the COVID-19 disease while highlighting the benefits for individuals’ receiving COVID-19 vaccines (French et al., 2020). These scholars concluded that social media platforms should not only eliminate misinformation from their platforms, but should also correct it with accurate information from authoritative sources. Facebook can apply these scholars’ recommendations to its strategy to fight misinformation on its platform during the COVID-19 pandemic.

Students should read and use the literature review to add scholars’ perspectives and suggestions to their recommendations for Facebook. The instructor can direct students to read the literature review if they would like to acquire a strong background on the scholars’ consensus regarding the controversial topic of information filtering. Additionally, students can view the 2010 documentary, “The Social Network,” to
generally understand the background of Mark Zuckerberg, founder of Facebook, and learn how he created the Facebook platform. Students should also watch the 2020 documentary, “The Social Dilemma,” for a more in-depth analysis of how social media platforms can create targeted content for their users and how these platforms have wide access to users’ data and browsing histories. These films will assist students in gaining a broad perspective on the developing topic at hand.
### APPENDIX

#### Exhibit 1

*Student Grading Rubric*

<table>
<thead>
<tr>
<th>Identification of Main Subject Matter / Issues</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identities and comprehends limited issues in the case</td>
<td>Identifies and expands on some issues in the case</td>
<td>Identifies and explains most issues in the case</td>
<td>Identifies and explains in-depth all issues of the case</td>
<td></td>
</tr>
</tbody>
</table>

| Analyses of Case Findings / Issues | Incomplete analyses of case findings and primary issues | Basic analyses of some case findings and significant issues | Thorough analyses of most case findings and major issues | In-depth and insightful analyses of all case findings and related issues |

| Suggestions for Appropriate Solutions / Implementation Strategies | Minimal suggestions for implementation strategies and related solutions to issues presented in the case | Some simple suggestions for appropriate solutions to some issues in the case | Appropriate suggestions and proposed solutions to most major issues in the case | Well-reasoned and thoughtful suggestions to address all major issues presented in the case |

| Additional Supporting Information / Reference to Related Readings | No outside research or reference to related reading or supporting information | Minimal outside research and limited reference to supporting information | Good additional research and use of supporting materials and readings | Strong outside additional research and excellent use of materials related to the case analyses |

| Writing Mechanisms | Poor use of proper grammar, spelling, and punctuation in responses | Inconsistent use of proper grammar, spelling, and punctuation in responses | Proper grammar, spelling, and punctuation in responses, with few errors | Excellent grammar, spelling, and punctuation in responses, with no errors |
### Exhibit 1

**Student Answer Scoring Rubric**

<table>
<thead>
<tr>
<th>Question</th>
<th>Learning Objective / Central Issues</th>
<th>Relevant Information</th>
<th>Proposed Solutions / Concerns</th>
</tr>
</thead>
<tbody>
<tr>
<td>What is Facebook currently doing to proactively and retroactively reduce online misinformation spreads on its platform?</td>
<td>Facebook’s current policies include: minimizing economic incentives, building new products, and helping people make informed decisions. Facebook is currently working to develop new AI technologies to improve its filtering technologies.</td>
<td>Facebook should continue to develop its AI technologies to improve Facebook’s post detection and filtering capabilities under the leadership of Mike Schroepfer. Facebook is still struggling to establish a clear position on addressing misinformation spreads, so it must closely define its objectives and actions in post filtering, especially with controversial issues, such as the COVID-19 pandemic.</td>
<td></td>
</tr>
<tr>
<td>How did Facebook’s current policies for removing misinformation from its platform and assess how its current information filtering policies align with these goals?</td>
<td>Following the outbreak of the COVID-19 pandemic, Facebook decreased its tolerance for misinformation and increased its efforts to flag and remove inaccurate information. Facebook also worked with third-party authoritative sources, such as the Center for Disease Control (CDC), to provide credible health-related information to correct online misinformation and disinformation.</td>
<td>Facebook must continue to use third-party authoritative sources to proactively introduce accurate information and retroactively present information corrections. Facebook must work to find an even balance between filtering information and remaining a source for the open discussions of information and users’ opinions.</td>
<td></td>
</tr>
<tr>
<td>How does misinformation negatively impact Facebook’s brand image, and how does filtering false or harmful information help Facebook’s brand equity?</td>
<td>Misinformation negatively impacts Facebook’s brand image because it reduces its information reliability. Online users and advertisers will not use the platform if they feel that the information will be inaccurate or misleading. Filtering harmful information on Facebook will help Facebook’s brand equity because users will feel more safe by using Facebook as a news source over its competitors that may lack any form of censorship.</td>
<td>Misinformation on the platform is also detrimental to Facebook’s customer lifetime value (CLV) because both advertisers and major news networks will stop purchasing promoted space on the platform if the amount of misleading and inaccurate information exceeds that of factually accurate information. Therefore, Facebook will suffer financially and lose its brand equity over other social media platforms if users perceive information on the platform to be unreliable.</td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>Understand the necessity of removing misinformation and assess the problems with filtering users’ content.</td>
<td>Facebook could face legal, ethical, and financial challenges by filtering information on its platform. The company must ensure that it is not limiting users’ rights to freedom of speech when filtering content. Facebook could also lose some of its customer equity by limiting users’ rights to freedom of speech because users might stop using the platform if it does not serve as an unregulated discussion board for opinions.</td>
<td>Facebook must be extremely cautious in differentiating between opinion pieces and harmful content that users post as purported factual information. This differentiation can be challenging for Facebook because individuals often interpret information subjectively.</td>
<td></td>
</tr>
<tr>
<td>How did the COVID-19 pandemic influence the spread of online misinformation throughout the social media industry and how did the pandemic impact the responses from various online companies?</td>
<td>The spread of misinformation on social media vastly increased concurrent with users’ use of social media platforms as means of gathering information and communicating with others. Social media companies faced the challenging situation of having to reduce misinformation and provide accurate information to users on a constantly evolving health topic. The platforms had to advance their technologies to keep up with the increased demand on social media and the limited supply of internal human content moderators.</td>
<td>Facebook made these decisions during a politically tumultuous time that left the country divided on many topics, including the COVID-19 vaccine uptake and distribution methods. Facebook must continue to take a proactive approach to filtering inaccurate information that can be harmful or misleading to its users. However, it must do so with caution so as to not lose large consumer segments from its current customer base.</td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>If you were Schroepfer, how would you handle removing misinformation on Facebook, and specifically, balance the risk of unintentionally removing accurate information against the risk of failing to identify inaccurate information? Would you need any additional information to make an informed decision?</td>
<td>Schroepfer should establish filtering policies that have minimal tolerance for misleading or inaccurate information. Schroepfer should also increase Facebook’s budget for upgraded technology for AI development and for additional human content moderators.</td>
<td>Schroepfer must protect Facebook users’ safety, even if doing so will come at a cost to the company’s profits and run the risk of erroneously deleting some users’ opinion pieces. To make this decision, Schroepfer should review detailed analyses of the company’s financials that includes the company’s daily user activity numbers during the pandemic and its weekly revenues following its new censoring policies and filtering initiatives.</td>
<td></td>
</tr>
</tbody>
</table>
REFERENCES


Facebook. (2011, August 16). *Case Study: Reaching Voters with Facebook Ads (Vote No on 8)*. https://www.facebook.com/unsupportedbrowser


CONCLUSION

This thesis has explored the impact of censorship on individuals and social media platforms, especially during health-related misinformation outbreaks. Specifically, this thesis has addressed the following question: *When health-related misinformation spreads on social media platforms, what are the appropriate responses for these platforms to reduce users’ exposure to false and misleading information, minimize potential adverse effects, and safeguard these platforms as conduits for free speech?*

This thesis uses a case study, a literature review, and an instructor’s manual to examine the thesis question. The case study discusses the history of Facebook and the role that Facebook’s Chief Technology Officer, Mike Schroepfer, played in developing strategies to combat misinformation on the platform.

The literature review explores leading scholars’ views on how to define misinformation, what factors encourage the spread of misinformation, and ways to efficiently correct online misinformation. The review compares scholars opinions and specifically applies to the COVID-19 pandemic and the subsequent misinformation outbreak.

Finally, the instructor’s manual explains how instructors can apply the case study to various courses and identifies key concepts that students should focus on when analyzing the case study. The instructor’s manual also details specific discussion questions for students to answer in class discussions.

Case Study

The case study follows the history of Facebook, from its founding in 2004 to its position during the 2020 to 2021 COVID-19 pandemic. The protagonist of the case study is Mike Schroepfer, Facebook’s Chief Technology Officer (CTO), who worked to increase and improve Facebook’s misinformation filtering technologies.

The case study explores Facebook’s response to misinformation on its platform and the corrective censorship policies it enacted. Throughout its history, the company had to continually revise its approach to misinformation outbreaks and faced specific challenges regarding political- and health-related misinformation.
The case study concludes at the outset of the 2020 COVID-19 pandemic and the health misinformation that proliferated online. This erroneous information included unverified origins of the virus, and inaccurate information about the treatments, cures, and vaccines. To determine how to reduce the impact of the misinformation outbreak, Schroepfer had to employ the most appropriate information filtering policies for Facebook’s platform.

**Literature Review**

The literature review explores scholarly journal articles that examine how online users create and spread misinformation and how social media companies correct it. The literature review presents scholars’ different explanations for the creation and spread of misinformation and offers their various solutions to correct this misinformation.

While the scholars disagreed in their specific proposed causes for the creation and spread of misinformation, they did agree that three specific factors encourage misinformation outbreaks: the human confirmation bias, information silos, and echo chamber effects. The scholars found that the social value users gain by spreading misinformation often contributes to users sharing and reposting the inaccurate information.

In terms of correcting misinformation, the scholars varied largely in their proposed solutions. While many scholars found that user-focused education programs would be beneficial to reduce misinformation spreads, they differed in what this education would look like. Additionally, some scholars focused on using theory-based and information-based information corrections, both proactively and retroactively to correct misinformation.

Finally, the literature review addresses scholars’ views regarding misinformation during the COVID-19 pandemic. While all but one of the scholars emphasized a need for credibility and authority in information sources, the scholars varied in their views about how authorities should present the information to social media users.
Instructor’s Manual

The instructor’s manual opens with an overview of the case study and follows with instructions and guiding questions for instructors to use when teaching the case. The case study explains Facebook’s concerns and complications in its attempts to filter information on its platform. The case follows Mike Schroepfer, the Chief Technology Officer of Facebook, as he explored ways to reduce online misinformation on Facebook.

The instructor’s manual recommends the case for fourth year undergraduate students who are studying marketing, finance, or management. The manual suggests that these students explore and analyze the case study in the beginning of the semester because it will give students a strong foundation in online marketing concepts. The instructor’s manual provides learning objectives for the case and gives instructors detailed discussion questions to further encourage in-class discussion about the learning objectives. The questions have sample answers from both a regular student and an outstanding student.

The instructor’s manual also recommends other media information sources, such as the 2010 documentary, “The Social Network,” and the 2020 documentary, “The Social Dilemma.”

Research Significance

As of 2021, the study of misinformation outbreaks on online platforms remained incomplete because many unknown variables contribute to its spread. Specifically, researchers have limited access to reliable information regarding the trends of misinformation spreads and their effects on individuals and businesses.
This thesis explores some of these trends in online misinformation outbreaks and focuses on health-related misinformation, specifically. This exploration will help scholars and health officials develop a clear plan for addressing, preventing, and eliminating health-related online misinformation, especially during the COVID-19 pandemic and future disease outbreaks. Overall, this thesis confirms scholars’ idea that information corrections and preventative filtering measures can address damaging online misinformation outbreaks. However, authorities must take these actions with careful consideration to avoid unnecessarily censoring opinion pieces from social media platforms.

**Recommendations for Future Research**

The literature review concludes with recommendations for future research. First, it recommends that scholars study the prevalence of misinformation and the trends in during misinformation outbreaks. This research will be especially important with health-related misinformation to control and reduce misinformation outbreaks that harm communities and vulnerable individuals. Lastly, the literature review recommends that scholars and health officials study studying the impacts of online misinformation on consumers and companies. This research would include the impact that information filtering has on consumers’ perceptions of a social media firm. Additionally, this research could analyze companies’ corporate social responsibility (CSR) initiatives and examine how those initiatives impact consumer revenues, customer lifetime value (CLV), and brand loyalty.

The novel COVID-19 viral pandemic that began in 2020 focused the world’s attention on the importance of providing accurate information to online users, reducing individuals’ fears during health crises, and garnering support for sound public health policies. Future research in the area will add to this store of knowledge on this important topic.