User's ability to detect fake news in online environments.

Aleeza N. Gardner
University of Louisville

Follow this and additional works at: https://ir.library.louisville.edu/etd
Part of the Mass Communication Commons, and the Social Media Commons

Recommended Citation
https://doi.org/10.18297/etd/3143

This Master's Thesis is brought to you for free and open access by ThinkIR: The University of Louisville's Institutional Repository. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of ThinkIR: The University of Louisville's Institutional Repository. This title appears here courtesy of the author, who has retained all other copyrights. For more information, please contact thinkir@louisville.edu.
USER’S ABILITY TO DETECT FAKE NEWS IN ONLINE ENVIRONMENTS

By

Aleeza N. Gardner
B.S., University of Louisville, 2017

A Thesis
Submitted to the Faculty of the
College of Arts and Sciences of the University of Louisville
in Partial Fulfillment of the Requirements
for the Degree of

Master of Arts in Communication

Department of Communication
University of Louisville
Louisville, Kentucky

May 2019
USER’S ABILITY TO DETECT FAKE NEWS IN ONLINE ENVIRONMENTS

By

Aleeza N. Gardner
B.S., University of Louisville, 2017

A Thesis Approved On

April 24, 2019

By the following Thesis Committee

Thesis Director (W Scott Sanders)

Second Committee Member (Lindsay Della)

Third Committee Member (Jason Immekus)
ABSTRACT

USER’S ABILITY TO DETECT FAKE NEWS IN ONLINE ENVIRONMENTS

Aleeza Gardner

April 24, 2019

Technology has drastically changed the way people consume news. Due to the rise of online spaces built for users to share content, there has been a rise in the spread of false information. Social media platforms, Facebook specifically, provide their users with personalized content based on their interests. The ease of spreading information online coupled with the anonymity of the Internet, creates a breeding ground for potential misuse of information. This study sought to examine how users view themselves and others were influenced by fake news on social media if the post is denoted as fake, or if the post confirms preexisting beliefs.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>SECTION</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>iii</td>
</tr>
<tr>
<td>List of Tables</td>
<td>vi</td>
</tr>
<tr>
<td><strong>CHAPTER</strong></td>
<td></td>
</tr>
<tr>
<td>I. INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>Internet Responsibility</td>
<td>1</td>
</tr>
<tr>
<td>Timeline and History</td>
<td>2</td>
</tr>
<tr>
<td>Statement of the Problem</td>
<td>4</td>
</tr>
<tr>
<td>Purpose of the Study</td>
<td>5</td>
</tr>
<tr>
<td>II. REVIEW OF LITERATURE</td>
<td>6</td>
</tr>
<tr>
<td>Cognitive Biases Effecting Evaluation of Facts</td>
<td>6</td>
</tr>
<tr>
<td>Astroturfing</td>
<td>9</td>
</tr>
<tr>
<td>Perceived Influence</td>
<td>12</td>
</tr>
<tr>
<td>III. METHODOLOGY</td>
<td>16</td>
</tr>
<tr>
<td>Stimuli</td>
<td>17</td>
</tr>
<tr>
<td>Procedures</td>
<td>19</td>
</tr>
<tr>
<td>Measures</td>
<td>19</td>
</tr>
<tr>
<td>Participants</td>
<td>23</td>
</tr>
<tr>
<td>Analysis</td>
<td>23</td>
</tr>
<tr>
<td>IV. Results</td>
<td>26</td>
</tr>
</tbody>
</table>
## LIST OF TABLES

<table>
<thead>
<tr>
<th>Tables</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Message Exposure Groups</td>
<td>16</td>
</tr>
<tr>
<td>2. F Statistic and Related Levels for Each Effect Tested on the Third-Person Effect</td>
<td>26</td>
</tr>
<tr>
<td>3. Group Means and Standard Deviations for Third-Person Effect</td>
<td>27</td>
</tr>
<tr>
<td>4. F Statistic and Related Levels for Each Effect Tested on Perceived Information Accuracy</td>
<td>29</td>
</tr>
<tr>
<td>5. Group Means and Standard Deviations of Concordant/Discordant</td>
<td>29</td>
</tr>
<tr>
<td>6. F Statistic and Related Levels for Each Effect Tested on Behavior Intent</td>
<td>30</td>
</tr>
<tr>
<td>7. Means and Standard Deviations of Behavioral Intent</td>
<td>31</td>
</tr>
</tbody>
</table>
CHAPTER I
INTRODUCTION

Internet Responsibility

The Internet has created an environment where users can stay connected to the world remotely and in real time. People are now able to shop, get news, stream movies and games, and connect with other users both locally and globally. The vast amount of freedom the Internet provides is paired with an equal amount of responsibility for users and websites alike. The Internet provides a space where news companies can share breaking news within minutes of an event; but this rush to publish quickly, comes with a responsibility to ensure that the information is accurate and unbiased. Similarly, retailers can access customers in remote areas, and users can provide feedback on purchased items. However, retailers must be aware that customers have increased agency to hold them accountable for bad behavior negative feedback that is visible to other consumers.

Social media platforms, such as Facebook, are not exempted from social responsibility on the Internet. Facebook seeks to give users a clean and concise presence to share information and content with other users. On the site, users can create and share their own content or re-share content from other accounts. When users use Facebook, content relevancy is determined through artificial intelligence with algorithms sorting information based on, among other things, previous interactions of the users. As a result of the algorithmically curated newsfeed, every user experience is unique to them where
the most pertinent content appears first, while less relevant content is lost or muted. In turn users only have a limited time to browse social media. Therefore, the action is justified as it enhances the user experience by providing relevant-to-them information at the top of their feed. This has the consequence of allowing social media platforms to act as gatekeepers.

Users have a responsibility to fact check and be conscious of the information they are consuming. Furthermore, any account sharing information in online spaces has an ethical responsibility to ensure that shared information is accurate. Despite this, given social media platforms have the ability to show and hide content from users, it is their responsibility to inform users about any potential threats to the validity information occurring on the website.

One specific example of the “responsibility debate” with respect to information verification on social media can be seen in the case of Facebook’s experience with Russian accounts during the run-up to the 2016 Presidential election.

**Timeline and History**

While false information is not a new concept, the Internet allows people to spread information without proper fact checking. Because of the ease, people can create and spread news on their own with few or no implications. With the rise of social media, there was also a rise in fake news being spread on the platforms. The fake news came from various accounts, but specifically, there were bots from Russia sending out information. Facebook was aware of Russian accounts as early as 2013 and informed the FBI of potential Russian presence and fake news on Facebook (Hudgens & Newcomb, 2017). Specifically, online bots (i.e. automated accounts) were sharing news stories that
contained only partial facts and were becoming more prevalent. By 2016, Politicians called for awareness to be brought to fake news because it was influencing opinion and spreading misinformation (Hudgens & Newcomb, 2017). However, Facebook did not start notifying users until late 2017 (Hudgens & Newcomb, 2017). Users were exposed to false or misleading information for years before they became aware of the choices made by Facebook to keep these misleading accounts from the public. During the 2016 United States Presidential Elections, social media platforms announced the Russian presence. The Russian accounts were sharing fake news stories intended to misinform readers (Hudgens & Newcomb, 2017). Although the harm seemed minimal, the effects were seen across the platforms with people believing and re-sharing the posts as though they were truth. At first it was thought the impact of this presence was minimal - less than 2% of the overall user population consisted of fake accounts - but as time progressed, more accounts were discovered, specifically Russian bot accounts (Hudgens & Newcomb, 2017).

Fake news is not a new concept, but the Internet allows the volume and reach of fake news to increase. Social media platforms generate revenue by giving businesses the ability to advertise to users. The advertisements are sponsored and can target specific groups of people at a relatively low cost. Targeting for advertisements can be done by demographic and psychographic information provided by the user or gathered by the user’s online habits. Russian intelligence programs exploited Facebook’s advertising tools to reach American voters. It is reported more than $100,000 in paid advertisements were for Russian troll accounts. It is also reported more than 126 million Facebook users
were exposed to Russian backed content, an increase in what was previously reported (Bump, 2018; Hudgens & Newcomb, 2017).

The content the Russian accounts were creating was about many different social issues occurring in the United States, but the most frequently occurring themes were politically charged. One can only assume that these politically-oriented messages were designed to influence, in some manner, the election process. Although social media platforms initially denied content on their websites had interfered in the 2016 election (Hudgens & Newcomb, 2017), in 2017 Facebook began to notify users of any contact they may have previously had with fake accounts. They also started testing ways of alerting users of potential compromised content. One method used was flagging stories other users noted as potentially fake. The flag approach had drawbacks and was criticized by academics because it could offend some users with deeply held beliefs and attitudes (Facebook Ditches Fake News Warning Flag, 2017). Users were also able to see if they had any interaction with fake accounts through the help center feature on Facebook (Guynn, 2017). These reactive responses to the Russian message bot issue, were seen by many as too little too late (Facebook Ditches Fake News Warning Flag, 2017).

Statement of the Problem

Reliance on the Internet and social media platforms for information and news has become common. Given the speed at which online information travels, the news media is under pressure to use social media to disperse accurate and timely information. However, the anonymous nature of the Internet allows governments, political parties, and even individuals to pose as legitimate news media and spread false or misleading information to further their own political or economic goals. Additionally, social media
platforms have chosen to take a light touch approach to avoid encumbering free speech, which begets the question of whether these passive reactions (e.g., flagging posts as originating from a questionable source) have any impact on user cognition, conviction or behavior.

**Purpose of the Study**

This study seeks to test whether a passive disclaimer warning on social media posts impacts users’ perceptions of the posts and intended to share the posts with others. Specifically, it will assess the influence of the disclaimer on three key message effects identified as important determinants of persuasion.
CHAPTER II
REVIEW OF LITERATURE

Technological advancements grant freedom to access information remotely and in real time. The Internet allows people to stay up to date with news around the world as it happens. One notable issue with the Internet as a platform for news is the spread of misleading or false information to others either knowingly or unknowingly. Although the spread of misinformation is an issue in offline contexts, the difference between the online and offline environments is the potential audience reached. Specifically, an online post can go viral worldwide within hours whereas historically it would take days or weeks for same information to spread via word-of-mouth or the press. Furthermore, the Internet and social media websites allow users to have mass amounts of information at their fingertips. Because of the large amount of information users encounter daily, several heuristics are employed to sort through the “facts” and determine which are indeed based in reality and which are untrue. The first is cognitive bias.

Cognitive Biases Effecting Evaluations of Facts

Confirmation bias, an aspect of cognitive bias, is a deviation from rational reasoning based on preferences for information that affirm one’s worldview. A result of this is that people seek out confirming information and discount or avoid disconfirming information (Winter, Metzger, & Flanagin, 2016). Studies have shown that people avoid seeking information that contradicts their beliefs as a defense mechanism or simply to preserve their current worldview (Winter, Metzger, & Flanagin, 2016). In terms of social
media, people will choose to follow and/or interact with accounts that solidifying their pre-held notions and beliefs.

Continued effect is when people remember information as it is related to their preexisting beliefs, and in some cases distort the information. People may not question or seek alternative or counter arguments for information they already believe to be true (Knobloch-Westerwich & Meng, 2011). Continued effect is especially true in social media settings. Users have a limited amount of time to review information. Therefore, if a user sees information similar to what they already believe to be true, they will take it for truth without properly examining that information or seeking alternative information to contradict it.

Selective exposure is when social media users only see what other users who hold similar attitudes or beliefs as them post or share (Pearson & Knobloch-Westerwich, 2018). Although users theoretically have the potential to see information they would otherwise not be exposed to through their online network’s sharing and posting of information, in actuality they may only see information that is similar to their own beliefs (Eisend, 2017). However, people do not generally avoid information they do not agree with; they simply do not get exposed to it because of the natural tendency of people to cluster with others who have similar attitudes. While selective exposure is the result of active user choices in who and what they follow on social media, it is also occurs passively based on a person’s social circle where people tend to associate with others who hold similar views and attitudes (Winter, Metzger, & Flanagin, 2016). Thus, even if people do not actively avoid information that is not consistent with their own beliefs, they also do not seek alternative material. Knobloch-Westerwich and Meng (2011) argue that
selective exposure is a heuristic employed by people to protect their pre-held attitudes and beliefs to maintain a favorable self-image.

Confirmation bias, selective exposure, and the continued effect work in concert to skew social media users’ perception of what is true. First, selective exposure is much more likely to occur in online networks because the connections (e.g. being “friends” with someone) formed there are exclusively voluntary. For example, Facebook requires a user to “like” or “friend” someone to get that person’s or page’s posts on their newsfeed. This is a required step to see information shared by another user and results in the focal user’s newsfeed being tailored to what they want to see or have subscribed to see. Users will also see information their friends share regardless of the factuality or if they agree or disagree with that information. This may result in users being exposed to posts they may not have otherwise seen. Second, because online users must make judgements regarding a large number of messages that they see in a short timeframe (Antonopoulos, Veglis, Gardikiotis, Kotsakis, & Kalliris, 2015; Atwood, 1994), these judgements can be based on previous knowledge held about the topic of the message (Atwood, 1994). This is confirmation bias. Finally, when exposed to messages, people tend to remember messages in ways that confirm their preexisting beliefs (Atwood, 1994).

Selective exposure, confirmation bias, and the continued effect bias are becoming more problematic not only because of users having to choose what information they want to see and believe, but because of built in algorithms on social media websites. The algorithms used by social media websites sort the information shown to users based on several factors. One of these factors is the amount of “likes” or “shares” a post has and how many of a user’s friends have interacted with that post (Antonopoulos, Veglis,
Gardikiotis, Kotsakis, & Kalliris, 2015). Coupled with the algorithms sorting information and users selecting what source they want to receive their information from, selective exposure is a part of social media users have to accept in terms of using the platforms (Eisend, 2017). Another factor the algorithms use to sort information is who the user engages with on a regular basis on social media. For example, if a user often interacts with another user or a page, the algorithm will automatically sort that person or page to the top of the user’s news feed.

**Astroturfing**

Astroturfing complicates the process of users sifting through information on social media to determine what is and is not credible. Astroturfing is the “manipulation of public opinion, made possible by the anonymity of the Internet, which makes it difficult to know when commentary on social media is being manipulated by those with vested interests” (Peng, Detchon, & Choo, 2017). Originally, online astroturfing was predominantly found in marketing where businesses paid people to review their products online to deceive the general public and sway mass opinion. More recently, online astroturfing has found its way into all areas, including the political sphere. Propaganda, as distinct from astroturfing, is successful by targeting specific groups and solidifying already held notions and attitudes (Brennen, 2017). However, propaganda does not, by definition, involve hiding the identity of the source. Not all propaganda attempts to spread extreme or polarizing material. Some propaganda spread in the 2016 election contained small amounts of wrong information (Timberg, 2016).

Astroturfing is not confined to online spaces and is not a new concept (Lee, 2010). However, it has gained traction and awareness recently due to the ease and
anonymous nature of the Internet (Sisson, 2017). Specifically, the low cost and ease of creating online advertisements creates a perfect environment for astroturfers to target people and play on their views. Creating fake profiles on social media is fast and easy with very little repercussion (Peng, Detchon, & Choo, 2017). As a result, people or groups can create fake profiles and spread information without actually being tied to that information or being held accountable. For example, during the 2016 election, the ease of remotely creating social media accounts under pseudonyms meant that Russian agents were able to create accounts solely used to disperse false or hyper-partisan news with little to no repercussions.

The presence of astroturfing in social media also hinders users from being able to accurately assess source credibility. When fake profiles and pages are created on social media, the true source of the message can be impossible to find. Hidden sources create a host of problems for users because it forces them to evaluate the source credibility of information using heuristics as it is impossible to evaluate the expertise or benevolence of a source, two important factors that comprise source credibility, without knowing the identity of a source. One heuristics often employed by users to assess credibility on social media is the “bandwagon effect” or social proof. Thus, the behavior of others suggests how they should evaluate a source. A user may also choose to see how many followers the source has or how many “likes”, “shares”, or “comments” their posts have when viewing and determining to believe the information from that source (Castillo, Mendoza, & Poblete, 2013). In short, the lack of information regarding the original source means that users must apply their best judgement regarding the veracity of a message based on the information and account itself.
Consequences

Both selective exposure and astroturfing create an environment where users have to be suspicious of the information they are viewing. Although being critical consumer of online information is generally a good idea, cognitive biases deter users from seeking alternative sources or fact checking.

Social media users, who would typically rely on news sources to act as a gatekeeper to prevent the dissemination of false information, now have to take on the gatekeeper role on their own. Traditionally, trusted news sources would be selective in the information they cover. Journalists are held to a high standard and are internally checked to ensure the standard is being withheld. As the use of social media rises, and the ease of creating and dispersing information rises with it, users must be mindful of the information they are consuming and be cautious of believing a source upon first interaction with that source.

There are more consequences to astroturfing than simply deceiving users who see it as propaganda. Once people believe the information, they begin to spread it as truth to their own followers. Multiple accounts can share the same message solidifying the false information in the minds of users as credible. After viewing such information by fake accounts, people create a new normal for the type of informant they receive (Sisson, 2017). For example, if a user sees a post by a fake account that confirms what they already hold to be true, they will expect more information like that and reject information that contradicts that information. Although the concept of astroturfing may seem harmless, the attempted manipulation of mass opinion can create real problems for receivers, especially receivers of political astroturfing (Castillo, Mendoza, & Poblete,
2013). The problems include mistrust or simply believing information without fact checking of all news and information a user may receive on social media (Sisson, 2017).

**Perceived Influence**

**Third-Person Effect**

Third-person effect is defined by Davison as the overestimation of the influence of media on the attitudes of others. People tend to believe they are less influenced by the media than others who see the same message. (Davidson, 1983; Atwood, 1994). It is argued that all media is persuasive even if the intent of the message is not persuasive by nature (Davidson, 1983). Media messages from the news is important to consider when looking at third-person effect because it is meant to inform the public, but people must make their own judgment about the information they are receiving. When making judgements about a message people should first be aware there is the potential for all users to be influenced by the message. Evaluation of the message is generally based on a person’s bank of knowledge about a particular topic, it is then that a person decides to either accept or reject the message and comes to understand the potential conclusions others might draw from the same material (Atwood, 1994).

While there are several factors involved in third person effect, ego is one that influences a person’s perception of information. Ego involvement is “self-oriented and consists of self-esteem, self-identity, and moral values” (Obeidat, Xiao, Iyer, & Nicholson, 2017). The pre-held notions can create bias that a person uses to assess new information. The more a person cares about a specific topic, the higher their ego involvement; and the higher the ego involvement, the more difficult it is to change their attitudes and beliefs (Ramos, 2017). Highly ego-involved individuals were not only more
likely to reject messages that threatened their self-concept, but were also more likely to presume that those messages would exert a strong effect on neutral receivers to sway them from to side (Perloff, 1989). Therefore, it is important to know why a person chooses to accept a message because it will play a role in determining if they believe others will influence by a message. People are then more susceptible to messages if they have a vested interest or see the information as positive to their overall self-image and belief system (Buturoiu, Durach, Udrea, & Corbu, 2017). An example would be if someone received a message that supported already held beliefs and attitudes about their political party or religion. According to Buturoiu, Durach, Udrea, and Corbu (2017) the third-person effect is greater when media content is seen as negative. Therefore, if a person perceives a message as overall good, and it supports their pre-held notions, they are less likely to have third-person effect (Buturoiu, Durach, Udrea, & Corbu, 2017). People who are vested in messages are more likely to spend more time reading and making judgements about a message than those with little or no vested interest (Eisend, 2017).

Influences of the media can be seen through exposure of a message. Social media, Facebook specifically, is built to show users relevant material based on their interests. The uncontrolled filtering of messages creates selective exposure of which users may not be aware. The platform is built to appeal to each user and relay information to them quickly (Antonopoulos, Veglis, Gardikiotis, Kotsakis, & Kalliris, 2015). Users must make judgements regarding the messages they see in a short timeframe (Antonopoulos, Veglis, Gardikiotis, Kotsakis, & Kalliris, 2015; Atwood, 1994). These judgements can be based on previous knowledge held about the topic of the message (Atwood, 1994).
Rationale

Studies have shown that astroturfing is successful in persuading people or changing people’s attitudes and beliefs. Although a shift in mindset caused by astroturfing (while unethical) might have minimal consequences when done in marketing, political astroturfing has far more insidious effects. The presence of accounts intentionally spreading false or misleading information on social media during an election could sway public opinion on important matters including who to vote for and general knowledge about the opposing political parties. Therefore, this study sought to see if public opinion is swayed by information on social media websites.

H1: The presence of a disclaimer (flagged vs. unflagged post) and the accord of information in a Facebook post with participants’ political leanings (message support general political leaning vs. message is in conflict with general political leaning) will interact to affect perceptions by the participant of the likelihood that others attitudes would be affected by the information contained in the posts (the third-person effect).

The third person effect occurs in both online and offline spaces. First introduced by Davidson (1989), people believe others are more influenced by media coverage than themselves. This study seeks to see if this holds true when people are confronted with fake information on social media websites.

H2: The presence of a disclaimer (flagged vs. unflagged post) and the accord of information in a Facebook post with participants’ political leanings (message support general political leaning vs. message is in conflict with general political
leaning) will interact to affect participants’ perceptions of the accuracy of the information contained in the post (perceived information accuracy).

Finally, social media websites did several things to combat the spread of fake accounts and misleading information after the public became aware of the fake accounts on the platform. One of those efforts was flagging posts with a banner at the bottom warning users of potential inaccuracies. This study seeks to see if those efforts were effective or if they were merely an attempt by the social media platform to save face to their millions of users.

H3: The presence of a disclaimer (flagged vs. unflagged post) and the accord of information in a Facebook post with participants’ political leanings (message support general political leaning vs. message is in conflict with general political leaning) will interact to affect participants’ intention to share the post with others (behavioral intention).
CHAPTER III
METHODOLOGY

To test these three hypotheses, I designed a post-test only 2 x 2 factorial design where I created four different stimuli replicating the look and feel of Facebook posts to create the following message exposure conditions:

Table 1
Message Exposure Groups

<table>
<thead>
<tr>
<th>Concordant</th>
<th>Unflagged</th>
<th>Flagged</th>
</tr>
</thead>
<tbody>
<tr>
<td>Democrats and saw Black</td>
<td>Democrats and saw Black</td>
<td>Democrats and saw Black</td>
</tr>
<tr>
<td>Republicans and saw Blue</td>
<td>Republicans and saw Blue</td>
<td>Republicans and saw Blue</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Discordant</th>
<th>Unflagged</th>
<th>Flagged</th>
</tr>
</thead>
<tbody>
<tr>
<td>Democrats and saw Blue</td>
<td>Democrats and saw Blue</td>
<td>Democrats and saw Blue</td>
</tr>
<tr>
<td>Republicans and saw Black</td>
<td>Republicans and saw Black</td>
<td>Republicans and saw Black</td>
</tr>
<tr>
<td>Black Lives Matter post</td>
<td>Black Lives Matter post</td>
<td>Black Lives Matter post</td>
</tr>
</tbody>
</table>
IVs were: 1) concordant/discordant message (whether or not participant saw a message that aligned with his or her political leanings based on the Blue Lives Matter movement (BuLM) and the Black Lives Matter movement (BkLM)).

2) un/flagged message (whether or not participant saw a message that contained a disclaimer at the bottom stating that the message was “disputed by third party checkers”).

The study sought to determine how these two independent variables influence three dependent variables: perceived message effect on others, perceived information accuracy, and intention to share information with others. To test these effects, I designed four fictitious Facebook posts to satisfy the four conditions created by crossing the IVs. After describing the study using an approved IRB preamble for survey research, individuals who agreed to participate were randomly assigned into one of the four treatment conditions.

A 2 X 2 factorial design tested concordant/discordant and flagged/unflagged post in social media influences behavioral intent and perceived information accuracy when evaluating news in an online environment. Participants were randomly assigned one of four conditions mirroring the manipulation of social media notification protocol. While political affiliation is an independent variable, the concern of the study is not to determine how each party reacts to being exposed to fake news, but rather how the parties overall react to fake news.

Stimuli

The conditions shown to participants were mock Facebook posts directly reflecting posts made by misleading accounts on Facebook prior to the election during the 2016 presidential campaign. The mock Facebook posts consisted of images with text
headlines and post footer showing the number of times the post was shared, liked, and/or commented on by other users. They were created in Adobe Photoshop to resemble real world posts.

This study used a 2 X 2 experimental design to generate its stimuli. First, Facebook posts were created to reflect contrasting narratives about an encounter between police and a young black man from the perspective of Black Lives Matter activists and Blue Lives Matter supporters. The Black Lives Matter movement seeks to bring awareness and stop violence and oppression to African Americans. The Blue Lives movement strives to bring attention to the dangers and sacrifices of police officers. The Blue Lives movement was a direct response to the Black Lives Matter movement.

The Black Lives Matter and Blue Lives Matter posts mirrored each other with the page name, logo, photo, and text. The narrative of the Black Lives Matter post involved police approaching a young African American man while he was leaving his high school and treating him as a suspect when he reached to get his cell phone from his pocket. In contrast, the Blue Lives Matter narrative concerned a young African American teenager who when approached by police leaving his high school reached into his pocket for what was thought to be a weapon. Second, posts indicated whether they have been flagged or not as potentially misleading. Specifically, there were two versions of both the Black Lives Matter and Blue Lives matter posts. One version was a typical Facebook post without any indication that the contained information might be misleading. Another version of the post was identical with the same text, image, and headline, but included a flag at the bottom of the image that read, “This post has been disputed.” Facebook used similar flags at the bottom of their posts after the 2016 election to combat the spread of
fake news. The flag at the bottom of the post is intended to alert users that the information within the post may be incorrect. By notifying users of its potential inaccuracy, this study sought to examine whether using a disclaimer would influence user’s intention to share content or believe the contained information.

**Procedures**

The experiment was conducted online using Google Forms. A link was emailed or posted on Blackboard for students to follow. After clicking on the study link in the solicitation email, participants were randomly assigned to one of four conditions using a randomized URL redirect hosted by Google Scripts written in JavaScript. Students were directed to a Google Form where they were given instructions, shown a preamble consent a brief description of the purpose of the experiment. After viewing and consenting to participation via a preamble consent form, and asked to fill out a questionnaire to measure intent to share these posts. Students were presented with stimuli and asked questions to third person effect, perceived information accuracy, behavior intent, and demographic information. At the end of the questionnaire, students were given a link to a separate Google Forms and were asked to fill in their name and instructor information for the sole purpose of giving extra credit.

**Measures**

The study included two independent variables, three dependent variables, and a variety of demographic questions.

*Independent variables:* Two independent variables were created for use in this experiment. The first independent variable was researcher manipulated. I created two Facebook posts that contained the disclaimer statement at the bottom of the post
indicating that the facts of the post were disputed by third-party fact checkers. The second IV was concordant/discordant message. To determine if the participant saw a concordant or discordant message, a generic question to determine their political identification was asked, then the data were separated into Black Lives Matter and Blue Lives Matter and the political affiliation question was used to determine if the participant was a Democrat or Republican based on if they said they were very likely to vote for a Democratic candidate or not likely to vote for a Democratic candidate. More specifically, the concordant/discordant message groups were based on natural response variation from the following question: “If you know nothing about either candidate other than political affiliation, how likely would you be to vote for a Democrat?” The question was asked on a five-point Likert response scale from very unlikely to very likely. Participants were assumed to be identity as a Democrat if they gave a response of likely or very likely. Participants were assumed to be identified as Republican if they gave a response of neither, unlikely, or very unlikely. The concordant group was created by matching individuals identifying as a Democrat with the Black Lives Matter message stimulus and individuals identify as a republican with the Blue Lives Matter message. The discordant group was created by matching individuals identifying as a Democrat with the Blue Lives Matter message stimulus and individuals identifying as a Republican with Black Lives Matter message stimulus.

Dependent variables: The three dependent variables of interest in the study are discussed below. They include the perceptions by the participant of the likelihood that others attitudes would be affected by the information contained in the posts (the third-person effect); perceptions of the accuracy of the information contained in the posts (Perceived
information accuracy); and reported likelihood to share the posts with others (behavioral intention).

Third Person Effect. Two items were queried on a five-point Likert scale with response options ranging from strongly agree to strongly disagree was used to measure the third person effect. The measure for the Third person effect was adapted from McLeod, Eveland, and Nathanson’s (1997) scale for the third person effect. The two questions were asked to measure the difference in how much participants thought they were influenced by the conditions, verses how much they thought others would be influenced by the conditions. An example item read, “Overall how much do you think the attitudes of others would be influenced by news on social media?” The “effect” variable was created by subtracting and individuals’ response to the question about the level of influence others would experience from their personal perceived level of influence. Thus, negative variable scores should be interpreted as a respondent reporting that others will be more affect by the information than they believe themselves to be. In comparison, positive scores should be interpreted as a respondent feeling as though he or she would be more influenced by the information than others. The discrepancy variable had a mean score of 1.238 (SD = 1.52).

Perceived Information Accuracy. A five-point Likert scale ranging from strongly agree to strongly disagree was used to measure perceived information accuracy (accuracy, believability, and trustworthy). Perceived information accuracy was assessed using a modified scale from Appleman and Sundar (2016) to reflect information seen on a social networking website. Three questions were asked to measure how participants viewed the information they were seeing. An example item read, “This post was
accurate.” A composite scale score was created by averaging participants’ responses to the three items. The scale had a mean score of 5.98 (SD = 2.07, Cronbach’s Alpha = .67).

Behavioral intent. Behavioral intent was measured using items written by the researcher to measure if the participant would share the information with others. A five-point Likert scale was used with choices ranging from strongly disagree to strongly agree. Two questions were asked to measure if a participant is likely to share the information in the condition with others on a social media platform. An example item read, “I am likely to share this post with others.” The scale had a mean score of 1.97 (SD = 1.91, Cronbach’s Alpha = .845).

Demographic variables: To gain a better understanding of the distribution of participants around key demographic information, I asked the following questions for descriptive purposes:

Year in school. Specifically, I asked “What year in school are you?” with the following response options: Freshman, Sophomore, Junior, Senior.

Race. Specifically, “what do you identify as?” with the following response options: White, Black, Hispanic, Other.

Social media habits: to gain insight into how often a participant uses social media, questions were asked about usage and what accounts they followed. The following questions were asked:

Use Facebook. Specifically, “Do you actively use Facebook?” with the following response options: Very Rarely, Rarely, Sometimes, Frequently, Very Frequently.
Frequency of use. Specifically, “How many times a week do you log into your Facebook account?” with the following response options: 1-2 times a week, 3-4 times a week, 5-6 times a week, daily.

Account following: Specifically, “I follow national or local news station accounts on Facebook (NPR, Fox News, CNN).” With the following response options: strongly disagree, disagree, neither agree or disagree, agree, strongly agree.

Participants

I recruited undergraduate students at the University of Louisville (N = 179) from communication classes. Students could receive extra credit as an incentive to participate in the study. Although names were collected, they were kept separate from the data to ensure that the study results remained anonymous. Participants reported their race as White (70.4%), Black (9.5%), Hispanic (7.8%), Other (10.1%). Their year in school was reported as Freshmen (24%), Sophomore (27.4%), Junior (22.3%), Senior (20.7%). The questionnaire included basic questions regarding demographic/psychographic information at the end. Students were also asked about their social media habits. Participants reported using Facebook very rarely (20.7%), rarely (12.8%), sometimes (24.6%), often (25.7%), and very often (12.8%). They reported getting news from Facebook Strongly Disagree (36.9%), Disagree (34.1%), Neutral (12.8%), Agree (11.7%), and Strongly Agree (1.1%).

Analysis

Data analysis progressed in three phases. First, I engaged in data cleansing and preparation. Next, I tested my main hypotheses. And finally, I tested additional hypotheses on a subset of the data.
Data preparation: To begin the analyses of these data, I cleaned and checked the data for adherence to the underlying assumptions of a factorial analysis of variance. Data cleaning conducted in Microsoft Excel to find any outliers or missing data.

Tests of Main Hypotheses: Next a series of three 2 x 2 factorial ANOVAs were conducted to test each of the three main hypotheses.

H₁: The presence of a disclaimer (flagged vs. unflagged post) and the accord of information in a Facebook post with participants’ political leanings (message support general political leaning vs. message is in conflict with general political leaning) will interact to affect perceptions by the participant of the likelihood that others attitudes would be affected by the information contained in the posts (the third-person effect).

H₂: The presence of a disclaimer (flagged vs. unflagged post) and the accord of information in a Facebook post with participants’ political leanings (message support general political leaning vs. message is in conflict with general political leaning) will interact to affect participants’ perceptions of the accuracy of the information contained in the post (perceived information accuracy).

H₃: The presence of a disclaimer (flagged vs. unflagged post) and the accord of information in a Facebook post with participants’ political leanings (message support general political leaning vs. message is in conflict with general political leaning) will interact to affect participants’ intention to share the post with others (behavioral intention).
CHAPTER IV

RESULTS

Data Analysis

The data were analyzed using a 2 X 2 between groups factorial ANOVA. The independent variables are political affiliation, and identified and not identified posts; the dependent variables were user’s intent to share a post on Facebook and perception of fake news on social media. The data was cleaned using Microsoft Excel to check for any outliers through visual inspection. Cases with predominantly missing data was excluded from being analyzed. Data was analyzed at a significance level of p < .05 to reject the null. All analysis was run using SPSS.

Hypothesis Testing

Hypothesis one tested whether the presence of a disclaimer (flagged vs. unflagged post) and the accord of information in a Facebook post with participants’ political leanings (message support general political leaning vs. message is in conflict with general political leaning) will interact to affect perceptions by the participant of the likelihood that others attitudes would be affected by the information contained in the posts (the third-person effect).

The results of 2 X 2 factorial ANOVA indicated there is no interaction between flagged message content and concordance of message with political identity and the third person discrepancy effect variable $F (1, 171) = .559, p > .05$. Therefore, I failed to reject
the null hypothesis. There were, however, two main effects for each of the
independent variables (flagged: p=.04; concordance: p = .018).

Table 2
F Statistic and Related Levels for Each Effect Tested on The Third-Person Effect

<table>
<thead>
<tr>
<th></th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>FlaggedUnflagged *</td>
<td>.559</td>
<td>.456</td>
</tr>
<tr>
<td>Concordance</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FlaggedUnflagged</td>
<td>4.277</td>
<td>.040</td>
</tr>
<tr>
<td>Concordance</td>
<td>5.754</td>
<td>.018</td>
</tr>
</tbody>
</table>

Table 3
Group Means and Standard Deviations for Third-Person Effect

<table>
<thead>
<tr>
<th></th>
<th>Unflagged</th>
<th>Flagged</th>
<th>Main Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concordant</td>
<td>-.622 ± 1.14</td>
<td>-1.265 ± 1.753</td>
<td>-.989±1.55</td>
</tr>
<tr>
<td>Discordant</td>
<td>-1.341 ± 1.41</td>
<td>-1.643± 1.51</td>
<td>-1.488±1.46</td>
</tr>
<tr>
<td>Main Effect</td>
<td>-1.012± 1.34</td>
<td>-1.440±1.65</td>
<td></td>
</tr>
</tbody>
</table>

Figure 1 below demonstrates the significant main effect of the concordant and discordant
groups, as well as the main effect between the flagged and unflagged groups.
Hypothesis two tested whether the presence of a disclaimer (flagged vs. unflagged post) and the accord of information in a Facebook post with participants’ political leanings (message support general political leaning vs. message is in conflict with general political leaning) will interact to affect participants’ perceptions of the accuracy of the information contained in the post (perceived information accuracy).

The results of a 2 X 2 factorial ANOVA indicated social media users are aware fake news does influence other users, $F(1, 171) = .261$, $p > .05$. After data analysis, I failed to reject the null hypothesis. However, a main effect was found for concordant/discordant ($p = .000$). Table 3 below provides F statistics and p values for all of the effects tested for this hypothesis.
Table 4

F Statistic and Related Levels For Each Effect Tested on Perceived Information Accuracy

<table>
<thead>
<tr>
<th>Effect</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interaction between flagged/unflagged and concord/discordant</td>
<td>.261</td>
<td>.610</td>
</tr>
<tr>
<td>Concord/discordant</td>
<td>14.623</td>
<td>.000</td>
</tr>
<tr>
<td>Flagged/Unflagged</td>
<td>2.100</td>
<td>.149</td>
</tr>
</tbody>
</table>

Table 5.

Means and Standard Deviations of Perceived Information Accuracy

<table>
<thead>
<tr>
<th></th>
<th>Unflagged</th>
<th>Flagged</th>
<th>Main Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concordant</td>
<td>2.23 ± .738</td>
<td>2.136 ± .649</td>
<td>2.1780 ± .687</td>
</tr>
<tr>
<td>Discordant</td>
<td>1.8963 ± .693</td>
<td>1.698 ± .641</td>
<td>1.8008 ± .644</td>
</tr>
</tbody>
</table>

Figure 2

Provides the graph below demonstrates the significant main effect of the concordant and discordant groups.
Hypothesis three tested the presence of a disclaimer (flagged vs. unflagged post) and the accord of information in a Facebook post with participants’ political leanings (message support general political leaning vs. message is in conflict with general political leaning) will interact to affect participants’ intention to share the post with others (behavioral intention). The results of a 2 X 2 factorial ANOVA indicated that identifying fake or misleading news on social media does not change a user’s attitude toward that information, $F(1, 171) = .054$, $p > .05$. Therefore, I failed to reject the null hypothesis.

Table 6

F Statistic and Related Levels for Each Effect Tested on Behavior Intent

<table>
<thead>
<tr>
<th>Effect</th>
<th>$F$</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flagged vs. Unflagged *</td>
<td>.054</td>
<td>.817</td>
</tr>
<tr>
<td>Concordance</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Flagged vs. Unflagged</td>
<td>3.035</td>
<td>.083</td>
</tr>
<tr>
<td>Concordance</td>
<td>2.772</td>
<td>.098</td>
</tr>
</tbody>
</table>
Table 7
Means and Standard Deviations of Behavioral Intent

<table>
<thead>
<tr>
<th></th>
<th>Unflagged</th>
<th>Flagged</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concordant</td>
<td>1.218 ± 1.056</td>
<td>1.00 ± .995</td>
</tr>
<tr>
<td>Discordant</td>
<td>1.0113 ± .991</td>
<td>.726 ± .726</td>
</tr>
</tbody>
</table>

Figure 3 below demonstrates the lack of significant effects between the groups.
CHAPTER V
DISCUSSION

Overview

While some of the information distributed on the Internet is beneficial to users, fake or misleading information has become common and widespread (Peng, Detchon, & Choo, 2017). The evaluation of the credibility and accuracy of information on the Internet is the responsibility of both users and content creators. In the absence of information about the source, one heuristic users employ to assess the credibility of a source is how many people follow the account (Castillo, Mendoza, & Poblete, 2013). Perhaps, the most important implication of false or misleading information, is its ability to influence public opinion on controversial issues. Therefore, this study examined how false information about politically polarizing issues is evaluated by social media users with deeply engrained political ideologies.

These results solidify studies about confirming and disconfirming messages in that a significant main effect for concordance of the message was found for the third person effect and the perceived accuracy of information on social media platforms. There is a significant amount of coverage about fake news in the media and the denial it had any influence on the American people during the election (Popken, 2019). The results of this study show that people do feel they and/or others were influenced by what they have seen on social media. With respect to my first hypothesis test, respondents perceived that the message would have a similar effect on themselves as compared with others when the
message supported their political ideologies. But, when the message did not support their political ideologies (i.e., it was discordant with their political identity), they were more likely to report that others would be more affected than themselves. A similar result was noted for the effect of message concordance on respondents’ perceptions of the accuracy of the information presented. When the message supported their political identities, respondents were more likely to perceive the information in the message as credible as compared to when the message was in opposition to their political identities. Perhaps of more interesting note, however, the concordance of the message with political identity did not have a significant effect on intention to share the post, which suggests that sharing behavior may be predicated on factors other than perceived information accuracy and perceived influence. Although not tested in this study, additional factors such as the number of likes and shares a post has received should be included in future analyses.

The results of this study also show that when a disclaimer posted to social media message, its effect was minimal. The disclaimer did not impact perceived information accuracy, nor did it change respondents’ intentions to share the post. It did significantly impact respondents’ perceptions of how they thought others might be affected by the notice. That is, when the post was flagged as containing questionable fact, respondents thought that others would be more affected by this notation than would they. When the disclaimer was not present, the discrepancy between perceived message impact was less extreme; although respondents still tended to report that others would be more influenced by the information than they thought they would themselves. These findings indicate that Facebook’s disclaimer was essentially ineffective at combating the spread of fake news.
Participants in the study were aware of fake news, but still chose to be a part of the platform and use it, even on a weekly basis.

**Limitations of the Study**

There were some limitations to the study. First, students could have answered the questions in ways that they felt were socially desirable. Although the survey was anonymous, some students still could have not been completely truthful with how they answered the questions. Second, this study did not ask about participants’ political affiliations directly. While the experiment did ask about political identity, the question asked how likely a person was to vote for a generic democratic candidate which might not translate directly to official political affiliation. Additionally, this variable was created by grouping participants into either an “Identifies as a Democrat” or “Identifies as a Republican” group and did not leave room for individuals who might identify as an independent voter. Finally, participants were recruited from communication classes at the University of Louisville, which suggests that they might not have been highly engaged with political messages. Or, due to the focus of the class content, they therefore might be more familiar with fake news, biases, and the third-person effect than the general population. Lastly, there was a low level of internal consistency among the items measuring perceived information accuracy in this sample, which could have negatively affected the outcome of the statistical test that employed the composite perceived information accuracy variable as the dependent variable.

**Future Research**
With the increase of fake news on social media websites, people have to evaluate the credibility of news and information more than ever. Future studies could ask if there is a threshold of how much fake or inaccurate news users would tolerate on social media before abandoning the platform. Another question that could be asked is if different political parties perceive themselves as more vulnerable to fake news than other political parties. From this study, we have seen that each political party sees the other as being influenced by fake news, but how does each party see themselves in terms of vulnerability to fake news. Finally, future research should include additional indications of message popularity as moderators when testing the effect of disclaimers and message concordance on the outcome measures studied here, as indicators such as the number of likes, shares, and network source (e.g., which friend shared the post with you).

**Summary**

Thousands of fake news stories circulated social media involving politics and social movements. Information circulating on social media during the 2016 election made it difficult for some users to distinguish between factual and fake news. The efforts of fake accounts spreading misinformation easily and quickly created a mistrust in news coming from social media and forced users to use their best judgement when evaluating information. Despite knowing about the false information, users still engaged with and actively used social media during the peak of fake news.

From the data, it can be concluded that flagging posts on social media as potentially false or disputed information does not appear to have a significant impact on readers’ evaluations of the information. News on Facebook and other social media platforms is already not as credible as news from traditional journalistic sources because
of the ease and simplicity of creating and dispersing information. The findings of this study suggest that more needs to be done by the social media platform in addition to adding a disclaimer to questionable posts to help readers quickly and accurately identify fake news and misinformation.
REFERENCES


Rhine, R., & Severance, L. (1970). Ego-involvement, discrepancy, source credibility, and


Appendix A

Conditions

Yesterday a teenage boy was exiting his high school when cops accused him of recent crimes in the area. A standoff began when the boy reached into his pocket to get his cellphone to call his mother. Police took this as a threat and ordered the boy to raise his hands and stay still. They were able to approach him, but kept guns drawn. He was then placed under arrest for possession of a weapon on school property, resisting arrest, and threatening police officers.
Yesterday a teenage boy was exiting his high school when cops accused him of recent crimes in the area. A standoff began when the boy reached into his pocket to get his cellphone to call his mother. Police took this as a threat and ordered the boy to raise his hands and stay still. They were able to approach him, but kept guns drawn. He was then placed under arrest for possession of a weapon on school property, resisting arrest, and threatening police officers.
Yesterday a teenage boy was exiting his high school when cops approached him to ask about recent crimes in the area. A standoff began when the boy reached into his pocket to grab a weapon. Police took this as a threat and ordered the boy to raise his hands and stay still. They were eventually able to subdue him after he fought back. He was then placed under arrest for possession of a weapon on school property, resisting arrest, and threatening police officers.
Yesterday a teenage boy was exiting his high school when cops approached him to ask about recent crimes in the area. A standoff began when the boy reached into his pocket to grab a weapon. Police took this as a threat and ordered the boy to raise his hands and stay still. They were eventually able to subdue him after he fought back. He was then placed under arrest for possession of a weapon on school property, resisting arrest, and threatening police officers.
Appendix B

Dear Class,

You are being invited to participate in a research study by answering the attached survey about news on social media. The survey is expected to take around 10 minutes and as a thank you for participation your instructor may opt to offer extra credit to all participants. We hope that this research will enable us to see how people interact with news on social media.

Taking part in this study is voluntary. You do not have to answer any questions that make you uncomfortable. You may choose not to take part at all. If you decide to be in this study you may stop taking part at any time. If you decide not to be in this study or if you stop taking part at any time, you will not lose any benefits for which you may qualify.

If you choose to participate please follow the link below:

https://goo.gl/sVC8sY

You will receive one follow up email after the initial to remind you of the opportunity. If you have any questions about how your information will be used do not hesitate to ask either (Aleeza Gardner/Dr. Sanders) or myself.

Sincerely,

Dr. Scott Sanders, Principal Investigator
Assistant Professor
University of Louisville, Dept. of Communication
310 Strickler Hall
Louisville, KY 40292
scott.sanders@louisville.edu

Aleeza Gardner, Co-Principal Investigator
Graduate Student
University of Louisville, Dept. of Communication
310 Strickler Hall
Louisville, KY 40292
aleeza.gardner@louisville.edu
Appendix C

Questions:

Third Person Effect:
Overall how much do you think your attitudes are influenced by news on social media?
Overall how much do you think the attitudes of others would be influenced by news on social media?
(McLeod, Eveland, & Nathanson, 1997)

Perceived Information Accuracy (Message credibility):
The post was believable.
The post was accurate.
The post was authentic.
(Appleman & Sundar, 2016) - message credibility scale reflective measure of message credibility

Behavioral intent:
I am likely to share this post with my friends.
I am likely to follow this account to see future posts.

Demographic:
What year in school are you?
Do you actively use Facebook?
How many times a week do you log into your Facebook account?
I get my news from Facebook.
I follow national or local news station accounts on Facebook (NPR, Fox News, CNN)
I identify as: (race)
Is any member of your family employed in law enforcement?

Measures of Political Affiliation:
If you know nothing about either candidate other than political affiliation, how likely would you be to vote for a democrat?
Overall how much do you think the attitudes of Democratic voters would be influenced by news on social media?
Overall how much do you think the attitudes of Republicans voters would be influenced by news on social media?
CURRICULUM VITAE
Aleeza Gardner
1820 Westmoorland Way, Louisville, KY 40242
(502) 554-1673

EDUCATION
MA in Communication 2019
University of Louisville
BS in Communication 2017
University of Louisville

TEACHING
Public Speaking Instructor 2018-2019
Communication Department, University of Louisville
• Planned and taught course based on real world situations
• Integrated new technology into lectures and classroom discussions

Teaching Assistant 2017-2018
Department of Communication, University of Louisville
• Proctored exams
• Graded assignments and reported/maintained grades on Blackboard
• Guest lecturer in classes regarding social media and personal research interest

RESEARCH EXPERIENCE
Research Assistant 2017-2018
Department of Communication, University of Louisville
• Assistant to W. Scott Sanders and Itay Gabay, conducting primary and secondary research
• Assisted in data coding and analysis for research projects

CONFERENCE PRESENTATION

UNIVERSITY SERVICE
Graduate Student Counsel 2017-2019
University of Louisville
• Collaborated and voted on key issues and events affecting graduate students at the University of Louisville

45