

IDENTITY MANAGEMENT: TWITTER'S POLICY ON SYNTHETIC AND MANIPULATED MEDIA

Kenneth A.D. Palattao, Jr.
Creighton University

With the introduction of the internet to the public in the mid-1990's, society gained access to humanity's immense library of knowledge and information from the comfort and safety of home. Alongside the rapid advancements of the "World Wide Web" and consumer computing, organizations, small-businesses, and citizens alike utilized the power of the internet to disseminate information to a global audience. What soon followed was the development of social media organizations and companies that sought to provide instantaneous, safe, and free dissemination of information between individuals and organizations. As a result of the daily advancements in technology and constantly changing political and social norms, social media users are in a constant state of worry regarding the safety of their personal information and the messages broadcasted to them via other individuals and organizations.

In a world where our information could be leaked in a moment's notice, where technology is bridging the gap between imagination and reality, and where hate and falsified media can target specific individuals and communities on a global scale, internet users have a claim against social media organizations to be responsible promoters of internet conversation and to identify with common societal values. As such, many of the crises that social media organizations face are *user-generated*. The heightened emotional state that social media users experience due to potential threats to their person and identity from other users on social media, prompts social media organizations to enact new policy to reassure users amidst unexpected and uncertain user-generated issues. Social media platforms utilize their rhetorical prowess to promote new policies that emphasize the safety of the user, allow free and truthful global conversations, and establish a sense of shared identity between the user and social media organization. Among the public's newest concerns regarding social

media is the dissemination of synthetic and falsified media from deceitful organizations and individuals. In this paper, I discuss the rhetorical strategies that social media organizations, like Twitter, use to respond to user-generated crises, quell user fear, and create a shared identity between user and organization. Further, I explore the extent of Twitter's success in following the tenets of the identity rhetoric genre to create a synthetic and falsified media policy that bolsters their identity with their users.

When Users Create Crises

In order to understand the rhetorical strategies that social media organizations utilize to create a shared identity with its users when confronted with crises, it is important to evaluate the situation that surrounds a current social media crisis and analyze how social media organizations have responded to their users. One of many current social media crises that users have brought forward to social media platforms is the presence and growing dissemination of synthetic and falsified media. New technologies have made it easier than ever to create and publish false messages on social media platforms that can persuade and harm both individuals and whole communities. These messages include edited photos, videos, and audio, as well as deceiving or false distribution of text from a variety of individual or organizational sources that range in quality and veracity.¹

Specifically, in instances of crisis, the public tends to shift focus from traditional news outlets to social media and places credibility in sources that may be questionable in origin.² Examples of false media that have pushed social media organizations to make change include terrorists issuing false hazard calls to carry out secondary attacks, deepfakes to create fake political messages or pornographic content, and edited audio from Joe Biden's and Nancy Pelosi's political speeches.³ Terrorists have utilized social media to call for help in an

¹ Xialing Lin and Patric R Spence, "Identity on Social Networks as a Cue: Identity, Retweets, and Credibility," *Communication Studies* 69, no. 5 (2018): 2, doi:10.1080/10510974.2018.1489295.

² Ibid, 2.

³ Bruce R. Lindsay, "Social Media and Disasters: Current Uses, Future Options and Policy Considerations," *Journal of Current Issues in Media*

area outside of the initial attack to injure or kill first responders in an attempt to exacerbate the impacts of disaster. Artificial intelligence can now be used to map faces from celebrities and political figures onto adult film actors or alter a political figure's voice to distribute false messages. Falsified media has created a substantial problem for how the public assesses the credibility of social media organizations as truthful distributors of current events and has established an exigency for change in social media policy.⁴

The Exigency for Social Media Policy Change

As a result of user-generated crises, social media platforms are continually presented with new issues that need to be solved in a timely manner so individuals may speak freely, have their privacy protected, and receive truthful information through their platforms. The inability of social media organizations to quickly enact policy that appeals to the user base and allows the organization to maintain control results in unanticipated and potentially threatening backlash that could sway the user to other social media platforms and threaten the credibility of the organization. Organizations want to give people the power to speak freely, however, to what extent can people speak freely? Are individuals allowed to say or post anything they want regardless of its validity? What constitutes something that is untruthful versus comical and what makes something threatening or compromises the safety of those online?

Up to this point, the line between conduct and content has been difficult to differentiate.⁵ It is now evident that social media platforms need to answer these questions and implement new policy. Even so,

& *Telecommunications* 2, no. 4 (2010): 293, <http://search.ebscohost.com/cuhs1.creighton.edu/login.aspx?direct=true&db=ufh&AN=79919807&login.asp&site=ehost-live>.

⁴ Lin and Spence, *Identity, Retweets and Credibility*, 1.

⁵ Susanne Kopf, "Content Policies in Social Media Critical Discourse Studies: The Invisible Hand of Social Media Providers?" *Critical Approaches to Discourse Analysis Across Disciplines* 11, no. 1 (2019): 5, <http://search.ebscohost.com/cuhs1.creighton.edu/login.aspx?direct=true&db=ufh&AN=142017395&login.asp&site=ehost-live>.

organizational rhetors can frame the situation as an unanticipated and potentially enhancing opportunity. Organizations are presented with an opportunity to reaffirm their mission statement, reflect their ideological tendencies to differentiate themselves, and strengthen their credibility as organizations that listen to and identify with their users.⁶ The situation shows that social media policy is not perfect and is always in the process of reformulation.⁷

To whom are social media policies tailored?

To create a sense of shared identity in response to this crisis, social media organizations must speak to a multitude of audiences because social media is now an important element of interpersonal and mass communication.⁸ The primary audience in the situation is the functional audience composed of users. The users play an active role in the organization because they promote the platform's daily function by posting content, generating advertisement revenue, and influencing policy. The users took issue with falsified and synthetic media, called for change, and have a claim against social media organizations to find a way to fix the issue so they may feel safe using the platform.

Another audience is the stockholders as an enabling audience. Rhetoric and policy that allow the users to have an impact on the organization shows stockholders, who monetarily allow the organization to operate, that the organization acknowledges the crisis and can respond in a way that is recognized by the user. The comfortability of the user increases, helps generate revenue, and provides the stockholders a return on their investment. A successful rhetorical response has the potential to target other social media organizations as normative audiences. Timely and well-designed rhetoric allow social media organizations to reflect their mission statement and differentiate themselves from competitors.

⁶ Ibid, 9.

⁷ Ignacio Siles, "Inventing Twitter: An Iterative Approach to New Media Development," *International Journal of Communication* (19328036) 7, (2013): 2, <http://search.ebscohost.com.cuhsl.creighton.edu/login.aspx?direct=true&db=ufh&AN=99140339&login.asp&site=ehost-live>.

⁸ Kopf, "The Invisible Hand," 1.

Organizations can show what must be done not only to appeal to the user, but to promote a safe and healthy environment for global conversation at all levels. If an organization's policy is successful and is later adopted by other organizations, that platform's reputation and credibility will increase as a result of their pioneering of new means of social networking safety. The many audiences that social media organizations engage introduces a multitude of constraints and assets when developing rhetoric and new policy.

Obstacles and Advantages in Creating New Policy

The primary constraint for social media platforms is how they define manipulated and falsified media and what constitutes a threat to the safety of individuals and groups on social media. Social media organizations emphasize an environment where individuals can speak freely without barriers. However, in order to prioritize the safety of all users, there must be some limitations to what the users can say or post. Implementation of policy impacts social media data while changing and representing social norms.⁹ As such, it is difficult for social media organizations to develop policy that allows users to speak freely and truthfully, but also allow healthy discussion. Further, in developing a policy on manipulated media, organizations must clearly lay out all the details for how they are going to police manipulated media when it arises. Organizations must create clear guidelines for what constitutes manipulated media and adhere to those guidelines tightly even if it goes against the values of the organization or contradicts the values of certain groups. If the organization contradicts their own policy, the credibility of the organization and their value system will be at risk.¹⁰

Further, social media organizations need to consider the legal ramifications of their actions when developing new policy. Flagging, banning, or suspending media based on it being manipulated or untruthful could potentially lead to lawsuits or backlash towards the organization. Even so, there are potential assets for the organization in this situation. The situation develops into a crisis as a result of the users recognizing the importance of change in policy for their safety. By listening to the users and developing a successful policy, the

⁹ Ibid, 7.

¹⁰ Ibid, 9.

organization increases user satisfaction and safety while also promoting a greater sense of shared identity. The credibility gained from past organizational actions carries considerable weight in swaying the users while developing new and future policies.

Twitter as an Exemplar

To analyze how social media organizations respond to user-generated crises – specifically falsified and synthetic media – and create shared identity with their users, I decided to study Twitter. Twitter was founded in 2006 by cofounders Evan Williams and Noah Glass as a side project from their start-up podcast company Odeo.¹¹ After Apple’s release of iTunes, engineer Jack Dorsey proposed a platform where users could share their activities and locations amongst friends. The growing popularity of Short Message Service (SMS) served as the groundwork for Twitter’s prototype that was launched publicly in July of 2006.¹² Since its inception in the innovation-driven Silicon Valley, Twitter has presented itself as a user-oriented organization. Many of the important innovations that make Twitter the social media organization it is today are the result of the interactions between developers and users. Twitter created an environment where developers act as users and users can act as developers to participate in the development of features and policy while expressing shared values and cultural ethos.¹³ A few of Twitter’s important early innovations as a result of user feedback are the use of the @ symbol and the hashtag to link users together.¹⁴

Today, Twitter is an open social media platform that allows its users to blog, network, and collaborate with individuals from all around the world. Users can post 140-character messages or “tweets” through the Twitter app or website and like and retweet the messages of others.¹⁵ In addition to connecting Twitter’s 330 million users from

¹¹ Siles, “An Iterative Approach,” 5.

¹² Ibid, 7.

¹³ Ibid, 12.

¹⁴ Ibid, 13.

¹⁵ Virendra Singh Nirban and Aniruddh Agarwal, “Understanding Sentiment Polarity in User Generated Social Media Discourse,” *Amity Journal of Media & Communications Studies (AJMCS)* 5, no. 3 (2016): 172,

across the globe, Twitter plays an important role in the dissemination of organizational messages and current events. As such, Twitter's mission is "to give everyone the power to create and share ideas and information instantly without barriers... [to allow] a free and global conversation."¹⁶ To promote this global conversation, Twitter conducts itself as a growing and adapting organization that works to create a safe online experience for all users. In order to create this experience, Twitter constantly changes its user policies with feedback from users and academic sources to prevent circulation of hateful, deceiving, or false messages and content.

In response to the growing amount of falsified and deceitful media that is posted online and within their platform, Twitter has created policy and bolstered their identity with users via the Twitter Safety Account and blogposts from Twitter's security team. The Twitter Safety account posts updates to user policy and issues surveys that allow the user to impact policy. As the Safety account is an official account on the platform, users have direct access to verifiable and trustworthy information from Twitter through their app and website. This profile is the fastest and easiest link between the user and Twitter officials when issues arise and responses to the user are made. This profile utilizes typical 140-character tweets, gifs, videos, photos, surveys, links to external sites, and other symbolic imagery to update the public on policy and guide the user to actively participate in policy creation. These messages serve as easily consumable content that provide a distinct message to target audiences.¹⁷ With regards to the falsified media issue, tweets recognizing a need for change in policy and the desire for user feedback began on October 21, 2019.

The blogposts serve as a place for Twitter's security team to respond to user feedback and show how they plan to implement new policy. These blogposts are generally posted as links on the Twitter Safety account and delve into the details of Twitter's policy while

<http://search.ebscohost.com.cuhsl.creighton.edu/login.aspx?direct=true&db=ufh&AN=142059115&login.asp&site=ehost-live>.

¹⁶ "FAQ," Twitter, April 13, 2020, <https://investor.twitterinc.com/contact/faq/default.aspx>.

¹⁷ Alina Nechita (Vingan), "Online Brand Awareness. A Case-Study on Creating Associations and Attachment," *Journal of Media Research* 11, no. 2 (2018): 94, doi:10.24193/jmr.31.7.

reflecting the user and academic impact on new policy. With the falsified media issue, these posts have served to update the users on the progress of the policy, reflect survey results and statistics, and inform the public about previous policy that Twitter has made with user feedback in mind. The first blogpost was sent out on November 11, 2019 to ask users to complete a multilingual survey that would give Twitter a better idea of how the user would like to see changes in policy to combat the falsified media issue. These artifacts are particularly useful because they reflect Twitter's desire to identify with the concerns of the user amidst this crisis by asking them to actively participate in the creation of new policy and providing them with pertinent updates to instill a sense of security.

Twitter and Identity Rhetoric

To analyze how social media organizations create a sense of shared identity with their users, I analyzed Twitter's artifacts during the falsified and synthetic media crisis via generic criticism of their identity rhetoric. Generic criticism is established upon the principle that there are recurring rhetorical patterns present across different rhetorical situations that respond to like needs of different audiences.¹⁸

A rhetorical genre is composed of three distinct elements that unify to produce a unique artifact: situational requirements, substantive and stylistic characteristics, and the organizing principle. Situational requirements deal with how the situation is observed resulting in specific rhetorical responses. The substantive and stylistic characteristics are the elements the rhetor uses to respond to the situational requirements. Substantive characteristics make up the content of the rhetoric while the stylistic characteristics construct its form. The organizing principle is the central term that embodies the situational, substantive, and stylistic elements of a given genre.¹⁹

The situational requirements that constitute the use of identity rhetoric arise when an enhancing or threatening opportunity presents itself and the organization's actions are reflective of its core features and values. The situation promotes the organization to build or

¹⁸ Sonja K. Foss, *In Rhetorical Criticism: Exploration and Practice*, (Illinois, Waveland Press, 2009), 137.

¹⁹ *Ibid*, 137.

maintain their identity and set themselves apart from other organizations through their swift response in addition to consistent rhetorical messages. The substantive strategies of identity rhetoric regard Stuart Albert and David Whetten's three components of organizational identity which emphasize "the central character of the organization, the claimed distinctiveness of the organization from other organizations, and a consistency of these elements over time."²⁰ Through association and differentiation, organizations unify practices and values that the audience perceives as good with the organization and demonstrate that they are different from other organizations.²¹ The stylistic characteristics of identity artifacts allow organizations to create a shared identity with their audience and are primarily associated with an organization's use of branding through different visual and verbal elements.²² The organizing principle of identity rhetoric is the way in which an organization portrays itself to an audience. The organization does this by utilizing organizational image and reputation to create a sense of credibility, align the organization with societal values, and persuade the audience to accept the organization's rhetoric and handling of the situation.

Findings and Analysis of Twitter's Identity Rhetoric

In developing their response to synthetic and falsified media, Twitter embodies the central tenets of rhetoric from the identity genre in their user-oriented tweets and blogposts. Based on my identity-centered generic analysis of Twitter's artifacts, I found that Twitter appeals to its audiences, establishes a sense of shared identity, and quells user fear by strongly adhering to the situational requirements characteristic of an identity crisis. Further, Twitter exemplifies the substantive characteristics of association and differentiation, reflects the stylistic characteristic of branding, and conforms to the organizing principle of identity rhetoric by demonstrating its multi-faceted image and reputation.

²⁰ Mary F. Hoffman and Debra J. Ford, *Organizational Rhetoric Situations and Strategies*, (California, SAGE, 2010), 121.

²¹ *Ibid*, 124-125.

²² *Ibid*, 126.

Twitter and the Situational Requirements of the Identity Rhetoric Genre

In response to the synthetic and falsified media crisis, Twitter strongly adheres to the situational requirements of identity rhetoric by releasing consistent rhetorical messages, establishing firm deadlines, and calling upon their prior actions to bolster their development of policy and subdue user fear and uncertainty. In the handling and dissemination of messages throughout the synthetic and falsified media crisis, Twitter utilized their Twitter Safety profile as their primary means of communication with their diverse user base. This direct connection to the users allowed Twitter to broadcast rhetorical messages and updates quickly while also establishing a direct line of communication between user and developer. Twitter's usage of the Twitter Safety profile served to differentiate Twitter from other social media organizations and mitigate user fear by establishing a verifiable and easily accessible timeline of Twitter's actions throughout the process of policy development. In analyzing Twitter's response to this crisis, I found that their initial tweet regarding a need for change in policy was issued on October 21, 2019. This tweet was sent out to users very early in the synthetic and falsified media crisis and reflects to users that Twitter is a very competent and cognizant organization that is capable of quickly recognizing when a problem arises in the minds of their users. From this point onwards, users became aware that Twitter had established synthetic and falsified media as an issue and depending on Twitter's handling of the situation, users could either identify with or dissociate from the values of Twitter.

Dissemination of consistent rhetoric via their Twitter Safety profile aided Twitter in their development of policy and creation of shared identity throughout the user-generated crisis. From their initial tweet, Twitter established a steady timeline of updates and information for its audiences. After their initial tweet, Twitter sent out weekly and sometimes daily updates on the policy making process. These updates showed Twitter users that Twitter is adamant about tackling the issue as well as maintaining and growing their shared identity. In addition to regular messages, Twitter also imposed on itself strict deadlines for policy implementation. In their initial blogpost, Twitter announced that they would provide a final announcement at least thirty days before their new policy was intended to go into effect. This statement provided users another claim against Twitter to stay true to their word in releasing the policy

on time. Twitter nearly followed through on this claim and announced in a tweet on February 11, 2020 that they would release their new policy on March 5th. Twitter continued to make updates to the public throughout this twenty-three-day period and adhered to their deadline releasing the new policy on March 5th. Additionally, Twitter established and maintained their shared identity with their users by continuing to post updates to their policy after its implementation. An example of this is shown in a tweet from April 22, 2020 where Twitter states that they have “removed over 2,230 tweets containing misleading and potentially harmful content...[and their] automated systems have challenged more than 3.4 million accounts targeting manipulative discussions around COVID-19.”²³

By calling upon their prior successful uses of user feedback to develop social media policy, Twitter further reaffirmed its shared identity with its users throughout the duration of the synthetic media crisis. Within multiple posts, Twitter added links to old blogposts that relayed how user feedback was utilized to create new policy. By reflecting their positive prior actions to drive forward new policy, Twitter was able to minimize backlash from users when creating new policy. The elements of prior policy that Twitter utilized to promote their new feedback oriented synthetic and falsified media policy will be further analyzed in my discussion of Twitter’s reputation.

The Substantive Characteristics of Association and Differentiation

Twitter associates itself with societal values and goals while differentiating itself from other social media organizations during the synthetic and falsified media crisis by employing strong word choice and emphasizing the benefits of user feedback in order to create and develop a sense of shared identity with its users. In Twitter’s first tweet regarding the falsified media crisis, Twitter immediately associated itself with the value of active listening. Twitter stated, “We’re working on a new policy to address synthetic and manipulated media on Twitter – but first we want to hear from you.”²⁴ This tweet specifically shows the user that they are different from other social media organizations in that they want to listen to

²³ Twitter (@TwitterSafety), Twitter, April 26, 2020, <https://twitter.com/TwitterSafety>.

²⁴ Twitter, @TwitterSafety.

users before drafting policy and believe that the users should have a say in developing the platform that they use daily. Further, Twitter is representing themselves as a calm and level-headed organization that recognizes that there is a problem at hand but does not want to rush into policy development behind closed doors and develop a policy based solely on collected data and ideas from developers. Twitter further associates with the value of open and active listening by stating, "As we shape our approach to synthetic and manipulated media, we think it's critical to consider global perspectives."²⁵ This demonstrates that Twitter wants to be considerate to all voices in order to protect and display the interests of all their users while promoting a global discussion. Additionally, both tweets emphasize the assumed "we" not only in the sense of personifying the organization, but also in joining the users to the organization.

Association with the emotions of the user and the societal values of truth and responsibility help Twitter to further differentiate itself from other social media organizations. In a blogpost which contained the results of their survey, Twitter recognized "the threat that misleading altered media poses and [that users] want Twitter to do something about it."²⁶ This post reflects that Twitter is a responsible organization. From the results of the survey, Twitter decided that they will provide users with more information, label content that has been altered, remove content that is likely to cause harm, and will show enforcement action towards users who share false media. Twitter further demonstrates that it is a truthful organization and differentiates itself from other organizations by recognizing that they are not a perfect organization. Twitter stated that they "will make errors along the way" but they are committed to making things right and want to approach policy with democratic participation in mind.²⁷ This truthfulness further extends to Twitter's ability to monitor and police the complexities of hateful conduct on their platform. Twitter stated that, "We agree that these are difficult areas to get right, so we

²⁵ Twitter, @Twitter Safety.

²⁶ Yoel Roth and Ashita Achuthan, "Building Rules in Public: Our Approach to Synthetic & Manipulated Media," *Twitter* (blog), February 4, 2020, https://blog.twitter.com/en_us/topics/company/2020/new-approach-to-synthetic-and-manipulated-media.html.

²⁷ *Ibid.*

want to be thoughtful and effective as we expand this rule.”²⁸ To reflect greater responsibility for the content present on their platform Twitter has “developed a global working group of outside experts to help us think about how we should address dehumanizing speech.”²⁹

The Stylistic Characteristic of Branding

The variety of visual and verbal elements that Twitter utilizes in their tweets on the Twitter Safety profile and blogposts aid in establishing Twitter as a cohesive brand that identifies with the values of its users. In order to create a sense of shared identity between Twitter and the user, Twitter exemplifies many of the characteristic qualities of a strong brand as explained by Nechita (2018). Twitter, like many other experienced social media marketers, primarily utilized simple texts with strong messages to target their audience.³⁰ Due to the 140-character limit of tweets, Twitter had to establish short and cohesive messages that informed the audience of their intentions and progress during policy creation. Twitter reaffirmed their brand and identity with users by creating easily consumable content bolstered by strong word choice that emphasizes their association with societal values and distinctiveness from other social media platforms throughout the crisis. In addition to the precise selection of words, Twitter also used other means to effectively communicate their message.

Twitter employed short videos to demonstrate how their policy would be carried out on the platform. In these videos, Twitter used many visual elements to inspire confidence in the user and promote a positive perception of the organization and their new policy. In these videos, Twitter showed generalized examples of how media that was flagged or censored due to manipulation would look on the platform. Off to the side of these examples, Twitter added concise and informative text to explain the details of the policy. This concise text was several times larger than the text of the tweets in video. These statements inspire a sense of confidence and trust in Twitter because

²⁸ “Updating Our Rules against Hateful Conduct,” *Twitter* (blog), March 5, 2020, https://blog.twitter.com/en_us/topics/company/2019/hatefulconductupdate.html.

²⁹ Twitter, “Updating Our Rules.”

³⁰ Nechita, “Online Brand Awareness,” 94.

Twitter says that they are using outside sources to review the media and specifically emphasizes the words safety, reputable, and inside look when users observe manipulated media on the platform. Nechita (2018) expresses that “static or motion pictures can trigger a better content, compared to a text post, regardless of the level of the writing” which aids in establishing a connection between the user and the brand.³¹

To inspire a sense of security in their policy and in itself as an organization, Twitter utilizes colors and symbols in its messages. Twitter employs colors specifically in their blogposts to bring out certain feelings in the users. In analyzing the blogposts, I noticed that in posts where Twitter asked for feedback or reflected on what they learned from feedback, green was used in the heading and for words that linked to other pages. Often, green is associated with growth, harmony, safety, and to “go” or move forward. Twitter likely used green on these pages to reflect that they are a continually growing organization that desires to identify with its users through feedback and create a safe environment for global conversation. On the page where Twitter announced the details of their new policy, red was used in the heading and for words that linked to other pages. Red is often associated with energy, war, emotions, and warning. This may reflect that Twitter has identified with the emotions of the user and has put a lot of time and energy into a new policy to end the war of hateful and false media present on their platform.

Twitter also associates their logo with the symbol of a shield in its short videos. By using an impactful image like a shield that is “consistent with the [organization’s] message can help to create brand identity and establish social connection much faster and longer than other types of messages.”³² The inherent nature of a shield is to protect its user. Shields defend their users from harm, attacks from other individuals, and mitigate the user’s exposure to undesirable forces. Acting as a shield, Twitter works to defend the voices of its users while decreasing the amount of hateful or falsified content that its users face on the platform. This symbol shows users that Twitter is a safe platform that continually strives to protect the user through its policies.

³¹ Ibid.

³² Ibid.

The Organizing Principle of Identity Rhetoric: Image and Reputation

In order to promote positive change on their platform and link the identities of the user with Twitter, Twitter portrays itself as an open, transparent, and dynamic organization and calls upon its prior successes in developing policy based on user feedback. To create the image that Twitter is an organization open to all views and voices, Twitter stated in a blogpost that they are “an open service— [whose] rules reflect the voice of the people” as a result of their feedback-oriented policies.³³ Twitter portrays itself as a dynamic organization when they state that they are “always updating rules based on how online behaviors change.”³⁴ Here, Twitter reflects to their audiences that they recognize the power of social media in our ever-changing society. Social media, due to the global discussions it fosters daily, plays a tremendous role in the development and change of social and political norms. In demonstrating that they recognize their role in shaping present and future social conduct, Twitter effectively shows its users that “they continuously evolve to reflect the realities of the world we operate within” and share the identity and values of the user.³⁵

In order to speak to their credibility as an organization and lessen the emotional tension associated with the crisis, Twitter references its prior successful policies that were driven by user feedback. In a blogpost, Twitter stated that they have “called for public feedback previously because [they] want...[the] rules to reflect the voice of the people who use Twitter.”³⁶ In 2018, Twitter requested feedback regarding hateful conduct on the platform. In asking for feedback, Twitter wanted to obtain comments from “a wide range of perspectives and to hear directly from the different communities and cultures who use Twitter around the globe.”³⁷ In the two weeks that they left their survey open, Twitter collected comments from more

³³ Del Harvey, “Help Us Shape Our Approach to Synthetic and Manipulated Media,” *Twitter* (blog), November 11, 2019, https://blog.twitter.com/en_us/topics/company/2019/synthetic_manipulated_media_policy_feedback.html.

³⁴ Twitter, *@TwitterSafety*.

³⁵ Twitter, “Updating Our Rules.”

³⁶ Harvey, *Help Us Shape*.

³⁷ Twitter, “Updating Our Rules.”

than 8,000 people from 30 different countries. Twitter utilized this feedback to change their hateful conduct policy to make its language clearer, narrow down identifiable groups, and promote more consistent enforcement of the rules.³⁸

Contribution to Rhetorical Theory

From my analysis of Twitter's rhetoric as an exemplar of identity management during a crisis, it is evident that successfully creating and maintaining a shared identity is the result of a culmination of significant and interrelated factors. When organizations properly utilize association, differentiation, branding, image, and reputation, they can more easily handle the rhetorical situation while strengthening their shared identity with their audiences. Failure to appeal to any one of these rhetorical tenets can potentially decrease the credibility of the organization during the crisis, threaten the positive perspective of the organization from its audiences, and damage the once cohesive relationship between organization and consumer. Twitter's rhetoric during the falsified and synthetic media crisis demonstrates many of the subtle aspects of identity rhetoric and how they intertwine to establish organizational identity.

Organizations that effectively associate themselves with their audience and differentiate themselves from similar organizations during a crisis must emphasize open, consistent, and transparent messages to their audiences while demonstrating the importance of the user's feedback. This is especially critical if the organization plays a role in the development of social norms, communication, or dissemination of important information. Simple, open, and coherent messages not only create a brand that audiences want to identify with, but also demonstrate that an organization recognizes the problem, takes responsibility and fully discloses its actions in solving the issue, and will solve the situation in a timely manner. Commonplace symbols establish fast and long-lasting social connections between the organization and its audience. Alongside this, displaying past organizational success during similar crises and using feedback from users or consumers of the product or service, expresses that the organization identifies with its audience, cares about societal values, and wishes to develop and maintain a shared identity.

³⁸ Ibid.

Despite the rhetorical findings that I observed in my analysis of Twitter and their identity rhetoric during the falsified and synthetic media crisis, there are limitations to my analysis. Firstly, I only studied one social media organization during this crisis. The synthetic and falsified media crisis not only impacted Twitter, but also all other major social media organizations including Instagram and Facebook. A comparative analysis of the identity rhetoric and strategies employed by the other social media organizations during this crisis is necessary to better understand the identity rhetoric and identity management strategies of social media organizations. Second, in analyzing how Twitter creates a sense of shared identity amidst crises, I only analyzed the rhetoric of the falsified media crisis. An investigation of other identity crises that have impacted Twitter would further aid in understanding the rhetoric that social media organizations use and the rhetorical trends they follow during an identity crisis. Finally, to complete a more encompassing and thorough analysis of Twitter's identity rhetoric and explain how organizations create shared identity during crises, I should have also analyzed Twitter's crisis and policy rhetoric. This would have given more depth to my analysis, provided greater rhetorical concept connections, and would be a great starting point for future research.

Conclusion

Twitter's progression and implementation of their manipulated media policy is successful in fulfilling the characteristics of the identity rhetoric genre. Twitter like other social media platforms was presented with an unanticipated and potentially threatening situation in which they had to reconvey their sense of identity with their users in order to prevent backlash and maintain their public image and reputation. Through consistent messages and updates on their Twitter Safety account, Twitter was able to reflect a desire to listen to the unique voice of the user and respond to the crisis of manipulated media in an urgent and quick manner. Twitter reaffirmed its credibility to its audience by giving a timeline of when the policy was going to be implemented and delivered the policy on the expected date – which was announced through their Twitter Safety profile. Twitter differentiated itself from other social media platforms by expressing the importance of the unique voices of the users on its Twitter Safety profile, inviting the users to participate in a

multilingual survey, conveying the voices of the audience in its blogposts, and delivering a time-sensitive finalized policy that represented the users and trusted academic experts. Twitter's reputation as an organization that listens to its users was reaffirmed by instating another feedback-oriented policy that directly connects to the positive societal values held by the audience. By associating with their audience, Twitter created a shared identity with its users and furthered their brand as a safe and healthy social media platform in the presence of an unforeseen crisis.

BIBLIOGRAPHY

- "FAQ." Twitter, Inc. - Contact - FAQ. Accessed April 13, 2020.
<https://investor.twitterinc.com/contact/faq/default.aspx>.
- Foss, Sonja K. *In Rhetorical Criticism: Exploration and Practice*. Long Grove, IL: Waveland Press, 2009.
- Harvey, Del. "Help Us Shape Our Approach to Synthetic and Manipulated Media." Twitter. Twitter, November 11, 2019.
https://blog.twitter.com/en_us/topics/company/2019/synthetic_manipulated_media_policy_feedback.html.
- Hoffman, Mary F., and Debra J. Ford. *Organizational Rhetoric Situations and Strategies*. Thousand Oaks, CA: SAGE, 2010.
- Jones, Marc Owen. "Propaganda, Fake News, and Fake Trends: The Weaponization of Twitter Bots in the Gulf Crisis." *International Journal of Communication (19328036)* 13 (January 2019): 1389-1415.
<http://search.ebscohost.com.cuhsl.creighton.edu/login.aspx?direct=true&db=ufh&AN=139171770&login.asp&site=ehost-live>.
- Kopf, Susanne. "Content Policies in Social Media Critical Discourse Studies: The Invisible Hand of Social Media Providers?" *Critical Approaches to Discourse Analysis Across Disciplines* 11, no. 1 (January 2019): 1-19.
<http://search.ebscohost.com.cuhsl.creighton.edu/login.aspx?direct=true&db=ufh&AN=142017395&login.asp&site=ehost-live>.
- Lee, Eun-Ju, and Edson C. Tandoc. "When News Meets the Audience: How Audience Feedback Online Affects News Production and Consumption." *Human Communication Research* 43, no. 4 (October 2017): 436-49. doi:10.1111/hcre.12123.
- Lin, Xialing, and Patric R Spence. "Identity on Social Networks as a Cue: Identity, Retweets, and Credibility." *Communication Studies* 69, no. 5 (November 2018): 461-82. doi:10.1080/10510974.2018.1489295.
- Lindsay, Bruce R. "Social Media and Disasters: Current Uses, Future Options and Policy Considerations." *Journal of Current Issues in Media & Telecommunications* 2, no. 4 (August 10, 2010): 287-97.
<http://search.ebscohost.com.cuhsl.creighton.edu/login.aspx?direct=true&db=ufh&AN=79919807&login.asp&site=ehost-live>.
- Nechita, (Vingan), Alina. "Online Brand Awareness. A Case-Study on Creating Associations and Attachment." *Journal of Media Research* 11, no. 2 (July 2018): 91-111. doi:10.24193/jmr.31.7.

- Nirban, Virendra Singh, and Aniruddh Agarwal. "Understanding Sentiment Polarity in User Generated Social Media Discourse." *Amity Journal of Media & Communications Studies (AJMCS)* 5, no. 3 (March 2016): 172-79. <http://search.ebscohost.com/cuhs1.creighton.edu/login.aspx?direct=true&db=ufh&AN=142059115&login.asp&site=ehost-live>.
- Roth, Yoel, and Ashita Achuthan. "Building Rules in Public: Our Approach to Synthetic & Manipulated Media." Twitter. Twitter, February 4, 2020. https://blog.twitter.com/en_us/topics/company/2020/new-approach-to-synthetic-and-manipulated-media.html.
- Siles, Ignacio. "Inventing Twitter: An Iterative Approach to New Media Development." *International Journal of Communication (19328036)* 7 (January 2013): 2105-27. <http://search.ebscohost.com/cuhs1.creighton.edu/login.aspx?direct=true&db=ufh&AN=99140339&login.asp&site=ehost-live>.
- Twitter (@TwitterSafety). 2020. Twitter, April 26, 2020. <https://twitter.com/TwitterSafety>.
- "Updating Our Rules against Hateful Conduct." Twitter. Twitter, March 5, 2020. https://blog.twitter.com/en_us/topics/company/2019/hatefulconductupdate.html.