

Attacking ISIL on Twitter: Addressing Ethical Responsibility in the Weaponization of Social Media

Grace Callahan

North Carolina State University

Abstract

My research revolves around the lack of oversight of social media platforms and the resulting ease in which terrorist groups are able to spread their message, recruit sympathizers, and maintain their network of supporters. Twitter dominates the social media arena as the most popular social media platform for terrorist recruitment, radicalization, and coordination. Consequently, questions emerge related to the establishment of international Internet governance with the issues of distributing ethical responsibility, maintaining protection of the users' right to freedom of speech, and creating a universal definition of terrorism. To explore the role of nation-states, social media companies, and the public in combating terrorist propaganda posted on social media, this project will analyze the initiatives by each of these actors specifically within the context of ISIL and Twitter. Rather than focusing on the role of only one of these actors, I will compare the effectiveness of initiatives by these actors to discredit ISIL's self-projection in the media and the creation of regulations on social media content to offer unique insight into the identification of the best equipped force to serve as a counterbalance to ISIL propaganda. I argue that the most effective means of combating the spread of ISIL propaganda on Twitter involves a simultaneous joint effort by nation-states, social media companies, and the public to create a system of international oversight and regulation of social media abuse by terrorists. Exploring the role of Twitter in ISIL's terror campaign provides the necessary context to counter ISIL propaganda, a development with the potential to ultimately lead to the disruption of the terrorist organization's network.

Introduction: Combating the Spread of ISIL Propaganda on Twitter

In August 2015, Jaelyn Delshaun Young, 20, and Muhammad Oda Dakhlalla, 22, were charged with attempting to travel to Syria to join and provide material support to the Islamic State of Iraq and the Levant (ISIL) (Shoichet, 2015). Young and Dakhlalla engaged in online messaging using social media platforms with undercover FBI agents posing as members of ISIL, expressing a willingness to provide support to ISIL through fighting skills and medical aid (Fausset, 2015). As promising young students from Mississippi, with Dakhlalla starting graduate school and Young finishing up her degree, their family and friends were shocked at this revelation due to the lack of ISIL exposure in their physical environment (Shoichet, 2015). While the specific social media platforms remain unidentified in this case, social media has helped ISIL recruit an estimated 30,000 foreign fighters from over 100 countries (Shoichet, 2015) (Brooking & Singer, 2016). The story of Young and Dakhlalla is just one among many, and ISIL's slick social media campaign has proven too effective to ignore.

ISIL's use of Twitter and other social media platforms exposes the necessity in devising a strategy to counter the ease in which terrorist organization are able to produce and disseminate propaganda using the Internet. This paper will examine the role of nation-states, social media entities, and the public in combating the spread of ISIL propaganda through transnational collaboration efforts to monitor ISIL's social media activity and initiatives. Issues surrounding the implementation of regulatory measures on social media platforms are also explored, including delegating ethical responsibility among actors involved as well as the international community, defining the limits of content to be regulated, and balancing the protection of the right to freedom of speech and privacy by users with the need for international security. I argue that a collaborative effort between nation-states, social media companies, and the public in the establishment of international Internet governance presents the most effective means of combating abuse of social media platforms by terrorist organizations. As each of these actors are limited in their individual capability to thwart terrorist propaganda on social media, addressing this issue necessitates a combination of their strengths in order to establish an effective counterbalance to the dissemination of terrorist propaganda. Many scholars argue that an effective countermeasure has yet to be devised due to the continued growth of ISIL, highlighting how counterterrorism measures have lagged behind terrorists' use of social media platforms underlines the importance of an interdisciplinary approach. This research question is essential to address given the potential precedent that ISIL sets for other Islamic terrorist groups through its construction of an effective and innovative social media campaign for recruitment, fundraising, and networking.

This paper will explore how nation-states, social media companies, and the public are attempting to combat ISIL's abuse of social media by

examining differing perspectives by scholars as to which of these actors offers the most effective means of countering ISIL propaganda on Twitter. The issue of establishing international Internet governance will also be introduced, presenting the clash between protecting international security and an infringement on fundamental rights to freedom of speech and privacy. The debate surrounding the ethicality and feasibility related to the implementation of Internet regulation will be grounded in the context of ISIL and Twitter, with the overarching goal of devising an effective countermeasure to the spread of terrorist propaganda on social media platforms.

Background: A Snapshot of ISIL Twitter Activity

An estimated 90% of terrorism organized on the Internet is through the use of social media platforms, which are defined by the ability of its users to share information with others around the world through messages, pictures, videos, links, and other forms of communication (Wu, 2015, p. 283-288). As social media networks provide instantaneous access to information and the ability to network with potential sympathizers, these platforms also serve as an invaluable tool for terrorists to launch propaganda campaigns, promote coordination and communication between its members, and attract the attention of potential recruits (Secara, 2015, p. 78). Terrorist organizations have begun to move toward primarily using social media platforms as opposed to more traditional media sources as a recruitment tool, explained through the increased efficiency of social media's broader network reach and its function as a two-way communication outlet (Weimann, 2014, p. 2-3). However, with social media platforms overtaking traditional media outlets in fueling the spread of terrorist propaganda, new strategies must be developed and countermeasures adjusted to align with this shift. In distinguishing terrorism from hate speech, it is important to note that hate speech is an expression of hate directed towards a group of people on the basis of a factor such as race, ethnicity, religion, or sexual orientation (Definition of Hate Speech). According to the U.S. State Department, terrorism is defined as premeditated violence that is politically motivated with the goal of achieving an intended effect on its audience (Whitaker, 2001). However, classifying social media content as either terrorism or hate speech can be difficult in terms of interpreting motives and eliminating the swaying influences of semantics or speculation (Brown, 2015). Overlap can also exist between content that is categorized as terrorism or hate speech, as shown through the presence of hate speech in ISIL propaganda on Twitter.

As a result of its anonymity, large number of users, lack of physical borders, and fast dissemination of information on a global scale, Twitter's popularity with terrorist organizations has surpassed that of other social media platforms such as Facebook, YouTube, and Instagram (Secara, 2015, p. 77) (Klausen, 2015, p. 1-2). Twitter's unparalleled importance for

terrorist organizations such as ISIL is illustrated by the potential reach of the 9,100 tweets that the platform's 554,750,000 users posted every second by mid-2013, a number that has only been increasing since 2013 (Weimann, 2014, p. 8). An estimated 43% of Twitter's users are between the ages of 18 and 34, providing a platform in which ISIL is able to target disaffected youth and attract them to their cause through concentrated propaganda efforts (Weimann, 2014, p. 8). With its free and user-friendly application, Twitter's appeal for terrorist groups to spread propaganda stems from the platform's ability to provide instantaneous access to information and distribution to a wide audience (Weimann, 2014, p. 3). Instead of content solely emanating from official ISIL accounts and leaders, Twitter allows supporters to add to ISIL's campaign by posting original content, retweeting official ISIL tweets and using hashtags which lends the appearance of a natural bottom-up movement of posts (Klausen, 2015, p. 17) (Keagle & Vitale, 2014, p. 17). In an examination of 59 Twitter accounts attributed to Western foreign fighters located in Syria between January and March 2014, Klausen (2015) suggests that individual and official ISIL accounts are too closely integrated to have been coincidental, highlighting the level of content control exercised by accounts located in the insurgency zone attributed to ISIL (p. 17-19). Klausen's (2015) purpose in making this claim is to demonstrate the extent to which ISIL's Twitter content is controlled, arguing that the appearance of a natural bottom-up movement of posts is actually tightly regulated by jihadist organizers and further suggesting the ability for nation-states, social media companies, and the public to discredit such propaganda using the same social media platform (p. 17). Klausen (2015) underlines the importance in slashing ISIL's appeal to potential recruits by drawing attention to the construction of ISIL propaganda on Twitter, identifying the terrorist group's seemingly natural and spontaneous media postings as manipulations of social media's features to inflate the appearance of support for ISIL's cause and their number of followers (p. 17-19). As a result, although no state recognizes the authority of ISIL, the group's social media presence and effectiveness of their targeted propaganda commands the attention of the world (Cooley et al., 2016, p. 13).

While the adaptation of social media platforms by terrorist groups to further their own agendas is not a new phenomenon, ISIL's use of social media as a weapon of war is arguably unprecedented (Keagle & Vitale, 2014, p. 5-6). Although ISIL pushes anti-Western propaganda and has a goal of establishing an Islamic caliphate similar to other terrorist organizations such as Al-Qaeda, ISIL distinguishes itself through its innovative manipulation of social media (Keagle & Vitale, 2014, p. 6-7). ISIL activity on Twitter displays a cunning employment of social media tactics to bolster and control their projected image, blurring the distinction between their real-life operations and carefully crafted propaganda (Brooking & Singer, 2016). While Al-Qaeda in the Arabian Peninsula

(AQAP) utilizes the Internet primarily through traditional media outlets and heavily relies on their fundraising functions, ISIL uses social media as a way to spread a campaign of fear (Secara, 2015, p. 80). ISIL has outperformed other terrorist groups through the sheer amount and impact of its produced media, utilizing a social media strategy that exaggerates and projects its power, esteemed status, and military successes (Cooley et al., 2016, p. 11-19) (Keagle & Vitale, 2014, p. 1). The implications of this strategy revolve around the inflation of ISIL's perceived image by the public and potential recruits, attracting a disproportionate level of both interest and fear in the terrorist group. In February 2014, ISIL's hashtag was used more than 10,000 times per day on Twitter (Berger, 2014). While Jabhat al-Nusra, ISIL's main competition in Syria, rakes in a similar number of supporters as ISIL on social media platforms, the terrorist group averaged only 2,500 to 5,000 posts per day that included its hashtag (Berger, 2014). The widespread popularity of social media platforms such as Twitter allows terrorist organizations to control territory on both the ground and the Internet, resulting in a newfound ease in ability to spread their ideology, gather support, and coordinate attacks around the world (Tadjdeh, 2015, p. 32).

ISIL's modern technological approach coupled with the use of strong force, intimidation, and financial resources has projected their appearance of power into a multidimensional global campaign (Keagle & Vitale, 2014, p. 4). In April 2014, ISIL launched a Twitter app called the Dawn of Glad Tidings which allowed ISIL to post tweets including links, photos and hashtags on the user's behalf (Berger, 2014). As ISIL closed in on Mosul, an estimated 40,000 propaganda tweets were recorded as having been posted by ISIL in a single day through the app launching the campaign #AllEyesOnISIS (Berger, 2014). With a population of 1.8 million, Mosul was engulfed in terror with the impending capture of the city heightened by circulating images of ISIL's barbarity as media outlets had no time to identify the real from the false reports of ISIL's momentum and brutality (Berger, 2014). As a result of this projection of ISIL's inflated sense of power coupled with crumbling leadership and morale, the Iraqi army numbering 25,000 surrendered to an ISIL force of 1,500 (Berger, 2014). In June 2014, Twitter abolished the app and consequently silenced thousands of accounts supporting ISIL (Berger & Morgan, 2015, p. 25). Based on a sample of 20,000 ISIL supporter accounts between September through December 2014, the number of Twitter accounts supporting ISIL is estimated to be at least 46,000 (Berger & Morgan, 2015, p. 2). In response to the termination of the Dawn of Glad Tidings, ISIL supporters shifted to using a larger number of applications under different services to ensure the continuation of the promotion of ISIL tweets and hashtags in the event that some were suspended (Berger & Morgan, 2015, p. 25). Such applications allow for automated posts to be published to Twitter accounts, eliminating the need for actual individual

users to create posts and creating the illusion of widespread popularity for terrorist organizations.

With a social media strategy of pushing a slew of graphic and grotesque images, ISIL's brutality is clearly projected onto its followers' feeds. However, ISIL's social media strategy also simultaneously emphasizes the normalcy of the group through the use of images depicting its members playing Call of Duty or eating Nutella instead of wielding a gun (Tadjeh, 2012, p. 32). The implications of such posts strives to emphasize a sense of normalcy in extremism (Klausen, 2015, p. 17-18), drawing potential recruits with the enticement of a life on the battlefield not dissimilar to their current lives at home. With the help of retweeting ISIL supporter accounts and Twitter apps designed to push posts and hashtags, ISIL ensures that its content goes viral. By hijacking innocuous hashtags such as #WorldCup2014 and #Brazil2014, ISIL propaganda is able to flood previously benign corners of the Web (Keagle & Vitale, 2014, p. 8). Users searching for either of these hashtags were faced with an onslaught of images showing a severed head with the caption "This is our ball... It is made of skin #WorldCup2014" (Keagle & Vitale, 2014, p. 8). While ISIL does garner significant organic support on social media platforms, it is important to recognize the role of its calculated social media campaign and use of apps to inflate this sense of support online, promote engagement and project strength (Berger, 2014). Although Twitter's policies on the boundaries of its content and use of the platform by users states the company's reserved right to remove graphic content or posts that promote terrorism, the actual execution of suspending accounts raises further questions such as distinguishing illegal content from objectionable content and terrorist motives from self-determination (Wu, 2015, p. 293-294). However, ISIL's inflation of its power militarily, politically, and informationally on social media to attract sympathizers underlines the need for a combatant strategy to ISIL's power projection that enlists participation from nation-states, social media platforms, and the public (Keagle & Vitale, 2014, p. 10).

Initiatives to Counter Terrorism on the Internet

In an attempt to combat the abuse of social media platforms by terrorist organizations, various nation-states and social media companies as well as the public have begun to execute initiatives in an attempt to combat the widespread and global dissemination of terrorist content. As the Internet is a shared global entity, abuses of social media platforms are often not neatly contained within nation-state borders or the realm of national law. Consequently, it is necessary to view the assignment of ethical responsibility in addressing terrorist propaganda on social media as a transnational issue of shared responsibility. In an attempt to determine the most efficient strategy to counter ISIL propaganda on Twitter, this section will analyze ongoing initiatives by nation-states, social media companies, and the public in reducing ISIL's appeal and ability to instill fear and

attract new recruits in their target audience. Through the presentation of scholarly critiques to each of these actors' abilities to reduce the attraction and dissemination of ISIL propaganda, the shortcomings of each are revealed and underline the necessity in a collaborative force in order to successfully combat ISIL both on the ground and on the Internet.

The Role of the Nation State

In the discussion surrounding the role of the nation-state in combating ISIL propaganda on Twitter, the U.S. emerges as the nation-state most often discussed in the literature as a strong potential counterbalance in creating its own media to discredit ISIL propaganda. The relationship between the U.S. government and Twitter is rooted in the fact that social media platforms are subject to existing laws of nation-states in which they operate (Kjuka, 2013) (Country Withheld Content). Although nation-states have the ability to request for removal of specific Twitter content on the basis that it violates their local law, Twitter has denied the majority of requests by nation-states (Kjuka, 2013) (Guidelines for Law Enforcement). As ISIL's social media strategy leans heavily on portraying the normalcy of the group, the State Department has led efforts in posting its own media to combat such misinformation in languages such as Arabic, Urdu, Punjabi, and Somali (Tadjdeh, 2015, p. 33). The Center for Strategic Counterterrorism Communications (CSCC) was established in 2010 under the State Department as an enterprise designed to counter the projected normalization of ISIL's radical ideology and violence online (Cottee, 2015). Holding the overarching motto that media is more than half the battle, CSCC boasts more than 50,000 "engagements" related to combating terrorist propaganda on social media in Arabic, Urdu, Somali, and English (Cottee, 2015). Such "engagements" are defined as including the production of counter-narrative videos created from a montage of graphic ISIL footage or tweets underlining the hypocrisy in ISIL's message (Cottee, 2015). For example, in April 2014 the CSCC attacked the trending hashtag of #accomplishmentsofISIS, producing sarcastic tweets using the hashtag such as "starving people of #Aleppo" and "destroying mosques in #Riqqah" in an attempt to undermine a false portrayal of ISIL's achievements (Cottee, 2015). However, ISIL propaganda enjoys a considerable advantage over CSCC counter-propaganda in the shock value of its posts, beheading videos that go viral in gross fascination subsequently steamroll over CSCC posts (Cottee, 2015).

Conducting a comparison between ISIL's own circulating propaganda and the countering social media campaign waged by the CSCC, Sorenson (2014) evaluated the effectiveness of the social media campaigns of both forces by examining the popularity of their posts on Twitter (p. 26). While a recruiting hashtag by an ISIL jihadi collected 32 favorites, a CSCC video depicting the brutality of ISIL in an effort to discourage potential recruits received zero favorites (Sorenson, 2014, p. 26). This gap in

popularity on social media between ISIL propaganda and countering posts by the CSCC suggests a disconnect between the amount of media produced by the CSCC with its actual effectiveness in countering ISIL propaganda as measured by its widespread dissemination. Although the CSCC video “Welcome to ISIS Land” which was posted in English and Arabic collected approximately 900,000 views, Fernandez (2015) highlights the question of whether such counterterrorism methods actually reach its target audience of potential ISIL recruits (p. 15-16). A central issue to consider in the evaluation of all ISIL counter-narrative efforts, this question exposes the danger in relying on counter-campaigns which depend solely on views as a correlation of its effectiveness statistically in discouraging ISIL recruits. Such implications could restrict the exploration of alternative countermeasures to ISIL propaganda given this apparent effectiveness in combating ISIL’s appeal assumed from the total number of views rather than the users that are viewing.

The emergence of the U.S. as a leading nation-state in addressing ISIL propaganda on Twitter mirrors its leadership in attacking ISIL on the ground. By October 2015, the U.S. had completed 5,473 airstrikes against ISIL as compared to the 1,574 conducted by the coalition including countries such as the U.K., Canada, France, Saudi Arabia, and Turkey (Fantz, 2015). Suggesting that the role of the U.S. should be reduced in such campaigns, Sorenson (2014) instead emphasizes the role of media outlets based in the Middle East and North Africa in taking the lead to disseminate credible information to combat ISIL propaganda (p. 34). Sorenson (2014) dismisses the potential for the U.S. to lead an effective information campaign in countering ISIL, citing the lack of credibility attributed to such initiatives by Muslims (p. 34). Although efforts by other nation-states to combat ISIL may be on a relatively smaller scale when compared to initiatives by the U.S., a gap in the literature exists with the discussion of the role of other nation-states similarly working to reduce the pull of ISIL's social media narrative. However, it is important to consider difficulties presented with the bureaucratic nature of nation-states in challenging ISIL on its social media strategy. The difference in procedural aspects that governmental agencies must follow when posting to social media in comparison to ISIL arguably hinders the ability of nation-states to combat ISIL propaganda (Tadjdeh, 2015, p. 32-33).

Twitter and the Responsibility of Social Media Companies

As a result of the expanding growth and popularity of social media platforms, social media companies are at the forefront of the discussion revolving around the abuse of their platforms by terrorist organizations. Given that terrorist propaganda is being pushed using social media as a platform for its dissemination, the most obvious countermeasure in disrupting the flow of the network appears to be through social media companies themselves. Social media entities such as Twitter have been widely critiqued for not adopting a more active stance against the posting

and dissemination of terrorist propaganda, often accused of helping to facilitate the expansion of terrorist sympathizers (Fernandez, 2015, p. 23). Although the obvious solution would be for social media platforms to remove offensive content and ban the users responsible, several issues emerge related to the feasibility of such a measure.

Previously describing itself as a “global town square” in which news is widely distributed and emanating from the public in real time, Twitter has a history of defending freedom of speech on its platform (Leetaru, 2016). However, this stance has enticed terrorist groups to favor Twitter and drawn criticism from governmental agencies towards the unchecked abuse taking place on the platform (Benner, 2016). In examining how social media companies address terrorist content on their platforms, it is necessary to recognize their dual role as platforms engaged in expanding the uninterrupted flow of information worldwide while simultaneously functioning as a company working to protect their brand image and economic interests (Busch & Shepard, 2014, p. 293). In Twitter’s choice to focus on outwardly marketing its benevolent role as a social media platform in contributing to the spread of information internationally, Busch and Shepard (2014) argue that Twitter is attempting to disguise its interests as a business through this self-projection as a neutral social service (p. 294). Since Twitter’s public policy emphasizes their advocacy for users’ rights to freedom of speech, privacy, and net neutrality (Twitter Public Policy), a regulation of content would indicate a retraction on these goals and potentially drive users to favor other social media platforms given that the popularity of the platform is tied to the level of trust and credibility attributed by its users (Busch & Shepard, 2014, p. 309). Busch and Shepard (2014) seek to reformulate the view of social media companies by underlining Twitter’s function as a commercial entity while also serving as a platform for civic engagement (p. 309). Engaging with Fernandez’s (2015) argument for a larger role for social media companies in enforcing regulatory measures to combat terrorist propaganda, Busch and Shepard help to frame the complexity of social media companies as actors in combating the abuse of their platforms by terrorist organizations given their dual role in protecting their brand and economic interests (p. 311). As a result, social media platforms such as Twitter are arguably incapable of sufficiently regulating their own platform alone given that they are reluctant to take any regulatory actions that might harm their platform’s image (Wu, 2015, p. 300-301).

As a result of intensifying pressure in the wake of the Paris and San Bernardino terrorist attacks, Twitter suspended an estimated 235,000 accounts between February and August 2016 that were linked to the promotion of terrorism (Koh, 2015) (An Update on Our Efforts to Combat Violent Extremism, 2016). These accounts were terminated by Twitter due to crossing the boundary of freedom of expression through their promotion of terrorism, explicitly classified as being in violation of Twitter’s policy regarding the acceptable boundaries of content (The Twitter Rules).

However, the effectiveness of the suspension of Twitter accounts is questionable as terrorist groups have been known to subvert these regulations as new accounts are created to replace the old (Weimann, 2014, p. 9-10). Suspending terrorist accounts on Twitter is arguably ineffective at slowing ISIL's momentum, given the other ways that users can find to access the content (Aukerman & Oh, 2013, p. 253) (Maggioni & Magri, 2015, p. 88). New accounts with similar user handles are created and quickly accumulate thousands of followers, highlighting the well-connected and content interdependent network of ISIL in which users do not have to actively search to find the newly created replacement accounts (Keagle & Vitale, 2014, p. 8).

To examine the effect of account suspensions on the terrorist network's overall performance, Berger and Morgan (2015) collected statistical data from a sample of 20,000 ISIL supporter accounts from September through December 2014 (p. 54). Before Twitter started suspending supporter accounts in September 2014, their study found an estimated 40,000 tweets that included the most popular ISIL hashtag of the group's name in Arabic each day (Berger & Morgan, 2015, p. 56). By February 2015, the number of tweets featuring this hashtag was reduced to about 5,000 per day (Berger & Morgan, 2015). Berger and Morgan (2015) argue that this refutes those who question the purpose in suspending Twitter accounts because of the continued availability of ISIL propaganda through replacement accounts (Berger & Morgan, 2015, p. 56). Since 2014, ISIL traffic on Twitter has decreased by 45% (Twitter, 2016), suggesting the effectiveness of account suspension in slowing ISIL's momentum on the platform. However, it is worth examining whether this decrease in ISIL traffic as a result of Twitter's suspension of related accounts eliminated this ISIL traffic or simply drove the supporters to favor another social media platform.

The restriction of terrorist content on social media also exposes potential ramifications on the ability of intelligence agencies to gather valuable information on current ISIL members, analyze tactics in recruitment, and identify potential recruits (Secara, 2015, p. 78). According to Donald Rumsfeld, the secretary of the U.S. Department of Defense, about 80% of the counterintelligence gathered on opponents is found through publicly accessible sources of information, including social media networks (Secara, 2015, p. 81). Berger and Morgan (2015) similarly underline the intelligence value in being able to examine ISIL supporter Twitter accounts, citing the large number of accounts that provide GPS coordinates within ISIL controlled territory despite warnings by ISIL to its followers to restrict this type of information (p. 54). However, Berger and Morgan (2015) argue that the Twitter accounts of ISIL supporters with smaller numbers of followers hold a greater intelligence value as a result of having a higher percentage of original content (p. 56). These smaller supporter accounts are also less likely to be suspended, with ISIL related accounts having higher numbers of followers and consequently stricter

oversight in content by ISIL more likely to be targeted in account suspension efforts (Berger & Morgan, 2015, p. 56). According to Berger and Morgan (2015), thousands of ISIL related accounts could be suspended without negatively impacting potential intelligence efforts (p. 56).

Bringing the Fight to the People

As Twitter is a platform based around user-generated content, ISIL's abuse of the platform does not have to be met without rebuttal. Instead, counter-narratives to ISIL propaganda produced by the public present another strategy to diminishing the terrorist organization's appeal. Since media promoting ISIL is aimed at mass recruiting their target audience of disaffected youth, contrasting stories by their peers utilizing the same social media platform can serve to combat ISIL propaganda without an infringement upon the right to freedom of speech. Although propaganda posted by ISIL on Twitter could remain on the platform, so could a denunciation of such content by users utilizing the same social media platform (Hanley, 2014, p. 1-2). An example of a social media campaign launched by the Active Change Foundation used the Twitter hashtag #NotInMyName, consisted of users seeking to counteract the perception of ISIL as representing the views of all Muslims (Hanley, 2014, p. 1-2). This social media campaign gained popularity with users seeking to combat the perception of ISIL activities as representative of the views of all Muslims, with tweets that included a hashtag criticizing ISIL and denouncing their extremist ideology in an attempt to dispel criticism that the Muslim community remains silent on issues of Islamist terror (Hanley, 2014, p. 2). Using social media platforms such as Twitter to combat propaganda by terrorist groups who have used the same platforms presents a notable alternative way to combat ISIL propaganda without the restriction of users' right to freedom of speech. Hanley (2014) develops this argument using the example of the condemnation of the beheading of American journalist James Foley by the Council on American-Islamic Relations (CAIR), which labeled this action as a violation of Islamic beliefs and universal norms regarding the treatment of journalists during conflict (Hanley, 2014, p. 1). A joint effort by CAIR and the Fiqh Council of North America published an open letter that gained 120 signatures from religious scholars and Muslim leaders, using Islamic terminology to refute any claims of religious justification to ISIL's actions (Hanley, 2014, p. 2). These examples of the Muslim community contesting the ideology and propaganda presented by ISIL are arguably underrepresented by Western mainstream media, with headlines instead reporting stories believed to gain more traction such as Israeli Prime Minister Binyamin Netanyahu's speech to the U.N. which declared all fanatical Islamists from ISIL to Hezbollah to be of the same fanatical creed (Hanley, 2014, p. 2). As a result, this skewed media coverage arguably undercuts the public's perception of its own power in the social media war against ISIL. With the

continued prevalence of an association of Islam with terrorism in the media, social media campaigns such as #NotInMyName appear to have little effect in swaying public perception of the relation between Islam and terrorist groups such as ISIL and could discourage future initiations of similar campaigns.

The importance of the role of Muslim-American citizens, community leaders, and clerics in refuting ISIL propaganda through the spread of accurate information on social media platforms cannot be stressed enough (Hanley, 2014). Imam Talib M. Shareef's request for media platforms to refer to ISIL as the "anti-Islamic State" underlines the need for a clear rejection of ISIL's ideology and actions by the Muslim community (Hanley, 2014). This example highlights social media's ability to serve a dual role in countering ISIL propaganda, underlining the importance of word choice in media and introducing the possibility of using "the anti-Islamic State" to help break down unhelpful associations between the views of all Muslims and ISIL's actions. Hanley's (2014) argument also underlines the importance of the public in producing media to discredit ISIL propaganda, specifically the role of American-Muslims (p. 1-3). Cooley, Stokes, and Gines (2016) similarly emphasize the power of civil society enabled through the function of social media to simultaneously serve as a counter-balance to ISIL propaganda (p. 11-18). Through a study analyzing the differences in ISIL's self-representation in its publication Dabiq and in the responding tweets by Arab users, the results indicated that Arab users perceive ISIL to be less powerful and prestigious than the group presents itself (Cooley et al., 2016, p. 19). Cooley et al.'s analysis of Arab tweets reacting to the publication and their perception of ISIL establish the public as a powerful actor in the discussion of countering social media abuse by terrorist organizations through the use of these same platforms. Given the study's findings of these lower perceptions by Arab Twitter users of ISIL's esteemed status, group solidarity, and success than ISIL's presentation of itself in Dabiq, the public's use of Twitter presents a challenge to ISIL's branding and crafted image (Cooley et al., 2016, p. 19-20). With Twitter's ability to serve as a forum in which the public is able to discuss, rebrand, and challenge ISIL propaganda efforts, the importance of the public's production of counter-narratives is clear in combating the widespread dissemination of terrorist propaganda on social media platforms.

To Regulate or Not to Regulate?

The sheer volume of terrorist content on social media inevitably evokes a discussion of whether to regulate the Internet for terrorist propaganda and remove offending content, as the lack of oversight into the use of social media platforms allows terrorist organizations to spread propaganda, expand their network of supporters, and connect with potential sympathizers. In addressing the issue of dissemination of terrorist content on social media platforms, a potential solution appears to be the

implementation of a regulatory system in which user-generated content is monitored and removed if found to be in violation of a set of proposed standards. However, several issues emerge surrounding the regulation of these platforms such as the lack of a universal definition of terrorism, difficulties in the identification of assigned responsibility in terms of international governance of social media platforms, and potential infringements on the right to freedom of expression and privacy. This section will examine the feasibility of establishing international Internet governance, exploring the possibility of an international definition of terrorism as well as scholarly opinions surrounding the multi-stakeholder model in Internet governance. The exploration of Internet regulation in this section will be grounded in its potential applicability in the context of ISIL and Twitter, with the existing literature used in an attempt to devise an effective countermeasure to the spread of terrorist propaganda on social media platforms.

The Possibility of International Internet Governance

The 9/11 attacks drew attention to the widespread access to information on the Internet, introducing the need for some form of Internet regulation to support anti-terrorism measures (Watney, 2007, p. 42-44). As a result, the Internet evolved from its initial loosely developed system of self-regulation to having a recognized potential for a legal regulation of its content (Watney, 2007, p. 42-44). However, a defining characteristic of social media is in its ability to transcend borders, presenting difficulties in establishing which organization bears the jurisdiction and responsibility of overseeing the regulation of its content (Wu, 2015, p. 284-292). Since the Internet is a shared global entity, its abuses in the form of aiding criminal or terrorist acts are rarely neatly contained within national borders (Watney, 2007, p. 41-43). As a result of these characteristics, the establishment of a comprehensive and international definition of terrorism is crucial in enabling oversight of terrorist content on the Internet (Wu, 2015, p. 293-294). However, the distinction between what separates illegal content from objectionable content and terrorist motives from self-determination is blurred, presenting difficulties in international collaboration in the governance of social media. The absence of an absolute definition shared by the international community as a collective underlines differences in how individual states define terrorism as well as conditions defining acceptable justification for violence (Wu, 2015, p. 304). To make Internet governance possible, a potential international definition for terrorism would have to be broad enough to receive a consensus from the international community, yet specific enough to be enforceable. Such a definition might be easier to gradually develop by the international community labeling specific acts as terrorist acts to serve as examples rather than attempting to construct an overarching definition to the satisfaction of each individual nation-state. Nonetheless, without international consensus on what constitutes acts of terrorism, regulation of

social media platforms spanning across territorial boundaries is unfeasible (Wu, 2015, p. 293-294).

The literature related to the potential establishment of international Internet governance often refers to a multi-stakeholder Internet governance, a model in which the public sector, the private sector, and civil society institutions have a role in the establishment of policies related to the Internet's management (Marsden, 2008, p. 115). Claiming that an effective form of Internet governance necessitates such a collaboration between nation-states, corporations, and the Internet user, Marsden (2008) uses the example of the Working Group on Internet Governance (WGIG) as a previous effective collaborative force tasked with the topic of issues surrounding Internet regulation (p. 116). Created by the United Nations in 2004, the WGIG was constructed as a multi-stakeholder group through its inclusion of representatives from nation-states, the private sector, and civil society stakeholders to help identify and clarify the issues surrounding Internet governance and the implementation of policy solutions (Marsden, 2008, p. 116). However, Marsden (2008) seeks to redefine an increased role for the consumers within the multi-stakeholder model, citing the billion global users of broadband Internet and the prominence of user-generated content as underlining the necessity in considering the responsibility of the consumer in the issue of Internet governance (p. 115-116).

In contrast, research by Purkkayastha and Bailey (2014) critiques the multi-stakeholder model as the most commonly portrayed method of Internet governance, highlighting the misrepresentation of equality between stakeholder roles as a key flaw in this approach (p. 112). Using the example of the multi-stakeholder model within the Internet Corporation for Assigned Names and Numbers (ICANN), Purkkayastha and Bailey (2014) illustrate an alternative reality of the multi-stakeholder approach in practice in relation to Internet regulation (p. 112). Purkkayastha and Bailey (2014) argue that the multi-stakeholder approach serves to disguise an Internet governance led by the U.S government with only the support of the private sector; Foreign governments serve only an advisory role through the Government Advisory Committee instead of as equals to the private sector or civil society, with the exception of the U.S. government as the overseer of ICANN (p. 112). Purkkayastha and Bailey (2014) also argue that the reduced role of governments in the multi-stakeholder model of ICANN is part of a U.S. strategy to keep other nation-states from taking control of the Internet, raising further questions as to the ability of the U.S. to serve as the leading force for addressing social media abuse by terrorist organizations (p. 112). Through their research, Purkkayastha and Bailey (2014) seek to illuminate flaws in the multi-stakeholder model to prove the ineffectiveness of this approach in its application to ICANN and the issue of Internet governance as a whole. The feasibility of a multi-stakeholder approach appears to be questionable if applied within the context of ISIL and Twitter, given the differing

interests yet equal footing of nation-states, social media companies, and the public as stakeholders.

International Internet governance and the regulation of social media platforms also presents a challenge as such oversight threatens to infringe upon the protection of freedom of speech and privacy. The removal of content from social media platforms can be interpreted as a violation of the right of users' to express their opinions and ideas without censorship. However, Watney (2007) underlines the various approaches to regulation of Internet content, using the example of legislation in place to specifically combat the availability of child pornography on the Internet, to distinguish between the general censorship of information as is practiced in China, Iran, and Saudi Arabia (p. 47-52). Similarly, Marsden (2008) uses the example of "Cleanfeed," a content blocking system invented by British Telecom, which uses the Internet Watch Foundation as a self-regulatory identifier of child pornography (p. 123). Marsden (2008) claims that the effectiveness of "Cleanfeed" is supported by the consensus of stakeholders in condemning child pornography, demonstrating the ability of stakeholders to transcend conflicting interests when working together to combat an urgent issue (p. 123). While the model presented by "Cleanfeed" could potentially be applied to the regulation of ISIL propaganda by international governance, it also underlines the question of whether a international collective definition of terrorism can be agreed upon in order to regulate this content.

Conclusion

Tackling the abuse of social media by terrorist groups presents a multifaceted challenge that must be addressed by the international community with the utmost urgency. Given the nature of social media platforms such as Facebook and Twitter in enabling the instantaneous and unrestricted access to information on the global scale, further challenges are presented in addressing the abuse of such platforms. While the U.S. maintains an active role in working to counter ISIL's narrative on social media, its effectiveness in reaching and dissuading its target audience from ISIL's influence is unclear. The blows to the perception of the U.S. in the Middle East due to its patterns of political, cultural, and military intrusion throughout the region arguably overcasts its attempts to provide counter-narratives by undermining its credibility. Social media entities such as Twitter struggle to balance the removal of terrorist content with the protection of their platforms' image as promoting unrestricted access to information, a reflection of their dual role as neutral social service as well as a commercial enterprise with economic interests. The effectiveness of account suspensions on Twitter is also questionable. Although Twitter removed 235,000 accounts related to ISIL (Koh, 2015), it is important to consider that users can return to Twitter under different usernames and resume posting. While ISIL's momentum may be temporarily slowed, social media companies must adopt a more aggressive stance toward the

regulation of terrorist content on their platforms. ISIL is winning the war of social media, and new countermeasures must be adopted to dissuade the masses of potential recruits from joining ISIL's ranks. Given the shortcomings of nation-states, social media companies, and the public in single-handedly stemming ISIL's disproportionate projection of power on social media, a collaboration is necessary between these actors. This research question is also important to address as ISIL will not be the last terrorist group to utilize social media to their advantage and the creation of a precedent as to the role of social media platforms could limit the influence and destruction of terrorist organizations in the future.

Following the 2016 U.S. presidential election, Twitter and other social media companies have faced increasing criticism for an influx in the prevalence of hate speech on their platforms. Twitter has recently issued a statement outlining new measures in an attempt to define the line between freedom of speech and abuse, unveiling new plans to increase the ability of users to hide unwanted content and report posts they consider to be abusive (Benner, 2016). As Twitter has arguably not developed a comprehensive response to handling the hate speech on its platform due to its overarching stance on the protection of freedom of speech, these measures indicate an increase in Twitter's acceptance of responsibility and willingness to take further action. With this crackdown by Twitter on hate speech in the context of a post-election U.S., it will be interesting to see if similar aggressive measures will be taken towards the regulation of terrorist content on the platform. If these increased measures in restricting the use of Twitter to disseminate hate speech are any indication, ISIL's social media war with Twitter at the forefront could soon encounter increasing roadblocks.

References

- An update on our efforts to counter violent extremism. (2016). Retrieved from <https://blog.twitter.com/2016/an-update-on-our-efforts-to-combat-violent-extremism>
- Aukerman, R., & Oh, J.Y. (2013). Freedom of speech and censorship in the Internet. *International Journal of Management and Information Systems*, 17(4), 251-256.
- Benner, K. (2016, Nov. 15). Twitter adds new ways to curb abuse and hate speech. *The New York Times*. Retrieved from <http://www.nytimes.com/2016/11/16/technology/twitter-adds-new-ways-to-curb-abuse-and-hate-speech.html?ribbon-ad-id=4&rref=technology&module=Ribbon&version=context®ion=Header&action=click&contentCollection=Technology&pctype=article>
- Berger, J. (2014, Jun. 16). How ISIS games Twitter. *The Atlantic*. Retrieved from <http://www.theatlantic.com/international/archive/2014/06/isis-iraq-twitter-social-media-strategy/372856/>
- Berger, J., & Morgan, J. (2015). The ISIS Twitter census: Defining and describing the population of ISIS supporters on Twitter. *The Brookings Project on U.S. Relations with the Islamic World*, 20, 1-65.
- Brooking, E., & Singer, P. (2016). War goes viral. *The Atlantic*. Retrieved from <http://www.theatlantic.com/magazine/archive/2016/11/war-goes-viral/501125/>
- Brown, T. (2015, June 20). Terrorism v hate crime: How U.S. courts decide. *BBC*. Retrieved from <http://www.bbc.com/news/world-us-canada-33205339>
- Busch, T., & Shepard, T. (2014). Doing well by doing good? Normative tensions underlying Twitter's corporate social responsibility ethos. *Convergence: The International Journal of Research into New Media Technologies*, 20(3), 293-315.
- Coe, P. (2015). The social media paradox: an intersection with freedom of expression and criminal law. *Information & Communications Technology Law*, 24(1), 16-40.
- Cooley, S., Stokes, E., & Gines, A. (2016). Battle of the brand: How Twitter users in the Arab world challenge ISIL. *Romanian Journal of Marketing*, 2, 11-23.
- Cottee, S. (2015, Mar. 2). Why it's so hard to stop ISIS propaganda. *The Atlantic*. Retrieved from <http://www.theatlantic.com/international/archive/2015/03/why-its-so-hard-to-stop-isis-propaganda/386216/>
- Country withheld content. (n.d.). Retrieved from <https://support.twitter.com/articles/20169222>
- Definition of hate speech. (n.d.). Retrieved from <http://dictionary.cambridge.org/us/dictionary/english/hate-speech>

- Fantz, A. (2015, Nov. 27). War on ISIS: Who's doing what? *CNN*. Retrieved from <http://www.cnn.com/2015/11/20/world/war-on-isis-whos-doing-what/>
- Fausset, R. (2015, Aug, 14). Young Mississippi couple linked to ISIS, perplexing all. *The New York Times*. Retrieved from <http://www.nytimes.com/2015/08/15/us/disbelief-in-mississippi-at-how-far-isis-message-can-travel.html>
- Fernandez, A. (2015). Here to stay and growing: Combating ISIS propaganda networks. *The Brookings Project on U.S. Relations with the Islamic World*, 1-27.
- Friis, S. (2015). Beyond anything we have ever seen: beheading videos and the visibility of violence in the war against ISIS. *International Affairs*, 91(4), 725-746.
- Guidelines for law enforcement. (n.d.). Retrieved from <https://support.twitter.com/articles/41949#18a>
- Hanley, D. (2014). American muslims condemn the "Anti-Islamic State" over and over. *The Washington Report on Middle East Affairs*, 33(9), 1-3.
- Keagle, J., & Vitale, H. (2014). A time to tweet, as well as a time to kill: ISIS's projection of power in Iraq and Syria. *Defense Horizons*, 77, 1-12.
- Kjuka, D. (2013). When terrorists take to social media. *The Atlantic*. Retrieved from <https://www.theatlantic.com/international/archive/2013/02/when-terrorists-take-to-social-media/273321/>
- Klausen, J. (2015). Tweeting the Jihad: Social media networks of Western foreign fighters in Syria and Iraq. *Studies in Conflict & Terrorism*, 38(1), 1-22.
- Koh, Y. (2016, Feb. 5). Twitter suspended 125,000 ISIS-related accounts in six months. *The Wall Street Journal*. Retrieved from <http://blogs.wsj.com/digits/2016/02/05/twitter-suspended-125000-isis-related-accounts-in-six-months/>
- Leetaru, K. (2015, Aug. 27). Who's doing the talking on Twitter? *The Atlantic*. Retrieved from <http://www.theatlantic.com/international/archive/2015/08/twitter-global-social-media/402415/>
- Liptak, A. (2008, June 11). Hate speech or free speech? What much of West bans in protected in U.S. *The New York Times*. Retrieved from <http://www.nytimes.com/2008/06/11/world/americas/11iht-hate.4.13645369.html>
- Maggioni, M., & Magri, P. (2015). Twitter and the Jihad: The communication strategy of ISIS. *The Italian Institute for International Political Studies*, 1-168.
- Marsden, C. (2008). Beyond Europe: The Internet, regulation, and multistakeholder governance- Representing the consumer interest? *Journal of Consumer Policy*, 31(1), 115-132.

- Mills, A. (2015). The law applicable to cross-border defamation on social media: whose law governs free speech in 'Facebookistan?' *Journal of Media Law*, 7(1), 1-35.
- Mullen, J. (2016, Nov. 23). Twitter accidentally suspends its own CEO's account. *CNN*. Retrieved from <http://money.cnn.com/2016/11/23/technology/twitter-jack-dorsey-ceo-founder-account-suspension/>
- Purkayastha, P. & Bailey, R. (2014). U.S. control of the Internet: Problems facing the movement to international governance. *Monthly Review*, 66(3), 103-127.
- Secara, D. (2015). The role of social networks in the work of terrorist groups, the case of ISIS and Al-Qaeda. *Research and Science Today*, 3, 77-83.
- Shoichet, C. (2015) Mississippi: Daughter of cop, son of imam accused of plan to join ISIS. *CNN*, Retrieved from <http://www.cnn.com/2015/08/12/us/mississippi-couple-isis/>
- Sorenson, D. (2014). Priming strategic communications: Countering the appeal of ISIS. *Parameters*, 44(3), 25-35.
- Tadjdeh, Y. (2015). U.S. government trying to counter ISIL "Twitter consensus." *National Defense*, 99(737), 32-33.
- Twitter. (2016, Feb. 5). Combating violent extremism. [Blog]. Retrieved from <https://blog.twitter.com/2016/combating-violent-extremism>
- Twitter. (2016, Aug. 18). An update on our efforts to combat violent extremism. [Blog]. Retrieved from <https://blog.twitter.com/2016/an-update-on-our-efforts-to-combat-violent-extremism>
- Twitter public policy. (n.d.). Retrieved from <https://about.twitter.com/company/policy>
- The Twitter rules. (n.d.). Retrieved from <https://support.twitter.com/articles/18311>
- Twitter usage/company facts. (n.d.). Retrieved from <https://about.twitter.com/company>
- Vogel, L. (2016). Why are doctors joining ISIS? *Canadian Medical Journal*, 188(3), 177-178.
- Watney, M. (2007). The evolution of Internet legal regulation in addressing crime and terrorism. *Journal of Digital Forensics, Security and Law*, 2(2), 41-56.
- Weimann, G. (2014). New terrorism and new media. *Commons Lab of the Woodrow Wilson International Center for Scholars*, Retrieved from <http://www.wilsoncenter.org/publication/new-terrorism-and-new-media>
- Whitaker, B. (2001, May 7). The definition of terrorism. *The Guardian*. Retrieved from <https://www.theguardian.com/world/2001/may/07/terrorism>
- Wu, P. (2015). Impossible to regulate? Social media, terrorists, and the role for the U.N. *Chicago Journal of International Law*, 16(1), 281-311.

