Spring 5-26-2023

Hacker, influencer, counter-culture spy: cyberspace actors' models of misinformation and counter-operations

Benjamin Kessell
DePaul University, ben.kessell@gmail.com

HACKER, INFLUENCER, COUNTER-CULTURE SPY: CYBERSPACE ACTORS' MODELS OF MISINFORMATION AND COUNTER-OPERATIONS

BY

BENJAMIN KESSELL

A THESIS SUBMITTED TO THE SCHOOL OF COMPUTING, COLLEGE OF COMPUTING AND DIGITAL MEDIA OF DEPAUL UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE IN CYBERSECURITY

DEPAUL UNIVERSITY
CHICAGO, ILLINOIS
2023

DePaul University
College of Computing and Digital Media

MS Thesis Verification

This thesis has been read and approved by the thesis committee below according to the requirements of the School of Computing graduate program and DePaul University.

Name: Benjamin Kessell

Title of dissertation: Hacker, Influencer, Counter-Culture Spy: Cyberspace Actors' Models of Misinformation and Counter-Operations.

Date of Dissertation Defense: 5/26/2023

Filipo Sharevski, PhD

Advisor*

Filipo Sharevski, PhD

1st Reader

Janine Spears, PhD

2nd Reader

Jacob Furst, PhD

3rd Reader

** A copy of this form has been signed, but may only be viewed after submission and approval of FERPA request letter.*

Hacker, Influencer, Counter-Culture Spy

Cyberspace Actors' Models of Misinformation and Counter-Operations.

Benjamin Kessell

Abstract

As misinformation continues to spread on social media, its residents have begun to fight back, independent of any platform. This organic resistance to the diffusion of misinformation is a clearly observable phenomenon with roots in Anonymous' distributed campaigns from the 2010s outwards. Hacker and information security communities are acting in defense of some of their favorite spaces, most notably, Twitter. Security researchers of all stripes use it for sharing indicators of compromise but, as the diffusion of misinformation becomes more problematic it becomes more difficult to find signals in the noise.

These actors' response to the issues at hand is polarizing, some convinced that political (counter) argumentation is a perpetual nuisance while others passionately attack the roots of misinformation, wherever they perceive it. In this study, researchers interview twenty-three cyberspace actors, highly active on Twitter & Discord, to better understand their mental models of misinformation, how they conceptualize counter operations and their thoughts on the efficacy of the tactics, tools and procedures involved therein.

Dedications

This work is dedicated to my partner, my mother, my family, and my community, who knew I would accomplish important things, even when I thought I couldn’t.

This work was made possible through the guidance of Dr. Sharevski, without whom I would be just another nerd, raging at the void on social media.

A very special thank you to all those who participated: I hope the wisdom you shared with me will go on to help build important things.

Contents

Definitions

1. *Misinformation*: Refers to false or inaccurate information that is disseminated regardless of intent to mislead.[55]

2. *Disinformation*: Refers to the dissemination of misinformation is deliberate so to mislead the information consumers[58] or cause material harm.

3. *Open-Source Intelligence (OSINT)*: Refers to intelligence generated by the collection and analysis of publicly available information with the purpose of answering a specific query.[16]

4. *Cyber-actors*: Refers to a person who is has engaged in Hacking, OSINT investigations, Cyber-Threat Intelligence, or Information Security.

5. *DISARM Framework*: Is the open-source, master framework for fighting disinformation through sharing data & analysis and coordinating effective action.[12]. In this work, it will be used to classify any tools, tactics, or procedures survey respondents.
   disclose.

6. *Counter-misinformation operations*: Refers to actions a Cyber-actor or group of Cyber-actors take to prevent, disable, or diminish the capacity of a misinformation actor or group of actors to diffuse misinformation into an Online Social Network.

Chapter 1: Introduction and Scope

1.1 Introduction

Much of the existing research on misinformation comes from a framework designed to analyze large-scale, ethnographic studies aimed at classifying or identifying misinformation in individual systems. In this study, we shine a light on smaller-scale interactions in social media. Where the average social media user's experience with misinformation is often transactional, ingesting or redistributing, hackers or security influencers interact on a relational level, seeing its presence as a threat to a free or democratized cyberspace. Traditionally hackers and security professionals discuss cyberspace as if it were their residence, seeing the intrusion of misinformation a corruption of the freedom they must spread information as they see fit.

Where the academic discourse focuses on automated detection and platform level responsibility, information security professionals and hackers alike see misinformation as an existential threat. Information security professionals may find their feeds increasingly becoming home to fear uncertainty and doubt, where some hackers perceive a dangerous infection spreading in a public space where they feel comfortable congregating. These groups exist on polar ends of a cyberspace actor spectrum but their public interactions with misinformation illustrate a more unified mindset on how to interact with it.

Several observable counter-misinformation campaigns have been well documented by academia and the mainstream media, while the proverbial boots-on-the-ground remain absent from those narratives. These operations are often painted with broad strokes, occasionally subject to the political orientation of the media reporting it. Phenomenological studies of the people directly involved or merely present for the sum of these operations are often narratives peppered with bylines from famous hackers like Aubrey "Kirtaner" Cottle, the self-professed founder of the hacktivist collective known as Anonymous; this study

seeks to understand the organic response to misinformation on an interpersonal, micro scale, to better understand the motivations, tactics and steps forward as they are understood by the individuals who willingly engage in occasionally dangerous counter operations.

In a democratic society protest is seen as a necessity for an unheard or ignored social group to reach out to the elite or those in power and force them to listen. Online we are beset on all sides by enthusiastic, constant dissent or argumentation and many of us accepted that as commonplace or just the cost of socializing online. The gradual dissolution between cyberspace and "real life" has made it easier than ever to become politically radicalized just by being present on social media. This study aims to understand how people who identify online spaces and activity as major parts of their "real life" conceptualize the tensions misinformation introduce to the systems and spaces they frequent. Where Facebook or traditional forums may be conceptualized as inconsequential because of their abstraction from physical space, cyber actors understand that online is very much also "real life." By asking these actors how they would or if they have participated in counter operations, we might better understand how the individual can participate in maintaining the healthy flow of information in the smaller communities and across platforms on social media.

Where models for content moderation are often postulated on by punditry and those who have been moderated, discussion around how individuals directly affect un-moderated content from the perspective of those individuals is found to be lacking by the researchers. By earning the trust of cyber-actors, discussing misinformation & counter-operations with them researchers hope to survey or forms with information that affirms a bias or affects the system in a manipulative way. In studying the phenomenology of misinformation as presented to hackers, activists, information security influencers and information activists at a micro-scale it is our hope that we can better equip both platform and individual to push back on an endless assault on our cognition and ability to reconcile our biases with the truth.

Where an individual without the skills these cyber actors possess may feel lost or defenseless against the torrent of misinformation, the experiences of the cyber actors can better inform future grassroots or community led efforts against dangerous misinformation operations. Where playbooks can

be created with frameworks like DISARM, this research aims to inform the reader as to the phenomena involved when one deploys such a framework as an individual or member of it distributed group. In doing so it is the researchers hope to humanize the effects of misinformation on social media, abstracting it from and macro scale quantitative analysis to better understand the human cost of witnessing the intrusion of misinformation in a beloved social space and what inspires some to fight back.

1.2 Statement of the Problem

When presented with the problem of how to combat misinformation on social media individuals often indicate aversion to interaction or deference to a platform's moderation capability. The average citizenry of a social media platform's user base may not feel inspired to push back on bias affirming misinformation content to block or do combat in the comments. Misinformation and profit incentives have created an era of social media where a war for your attention both on and off the platform is constantly raging. Whether at the hands of a toxic oppositional ideology, state actors or capitalist elites, the user is victim to these cognitive aggressions as soon as they log on and off did not long after they log off.

True to their roots in counterculture, hackers and their comrades willingly stand against the elites declaring unity, freedom of information and a democratic internet worthy of direct-action activism. In seeking the wisdom of outsiders, of counterculture, we attempt to better understand the emergent civic strife that misinformation and advocates for the freedom of information are observably embroiled in. Rather than understand grassroots movements as distributed functions within a system we attempt to understand the motivations of those affected by witness to or engaging in first party direct action against misinformation on social media.

1.3 Research Questions

RQ1 How do people who consider themselves as active operational participants in the cyberspace, e.g., hackers, hacktivists, information security professionals, etc. conceptualize misinformation?

RQ2 What are the tools, tactics, and procedures these "cyber-actors" employ in dealing with any form of falsehoods or questionable information in cyberspace?

RQ3 What is the best response against misinformation in the view of these "cyber-actors"?

1.4 Significance of the Problem

Where large-scale data collection offers insight into how to create controls to counteract the disinformation diffusion on social media; it ignores the power projection of the outsider onto the wider social system unless discussing state action journalistic viewpoint. Where research like DISARM or The Credibility Coalition aids in triaging misinformation as the information security professional might triage a cyber incident is a net gain for the overall integrity of the internet; the stories of outsider actors is presented herein to inform a human-centered cybersecurity model where professionals and activists alike advocate for themselves in cyberspace.

Where grassroots resistance may galvanize around an exceptional or outspoken members of the resistance, their success is often the result of their decentralization and ability to self-regulate. Media often romanticize or demonize these outlier actors, further spreading the word of their movement. With the convergence of the physical and digital spaces thanks to decentralized activism organizing over social media, the tendency is to refer to the ideology over the individuals involved. When analyzing activism to counteract misinformation on social media, mainstream media dehumanizes the effort preferring instead to reference hashtags and vernacular nomenclature.

Where the breach of a particular properties boundaries often referred to as criminal acts it is the transgression on these boundaries that force them to flex with civil society. The cyber-actors' interactions

online are often transgressive of the platform's desires for its user base. The lack of insight into these actors transgressive or defensive interactions in social spaces online, from an objective viewpoint, presents a unique barrier to developing methods for educating or inspiring users to advocate for their cognitive independence from the battle for their attention. Much of the media investigating or profiling famous hackers has inspired many professionals to pursue information security. So too does the absence of material profiling those who would push the boundaries of the individuals' conceptual responsibilities when presented with misinformation present a barrier to the creation of future generations moved to organize and oppose the diffusion of misinformation into their communities.

1.5 Statement of Scope

The researcher has observed actors from inside the VX-Underground, 420Chan, Anonymous, Telecomix and, InfoSec Twitter & Discord communities, identifying active users for interviews to capture the following:

1. The subject's first-party model of misinformation.
2. Any tactics observed or deployed by the subject.
3. Human-centered or outsider models of counter misinformation operations.

Collected surveys are analyzed for the following:

1. Classifiable sentiment using the Sharevski Et Al Folk model[46] as a primary classifier and the Wu Et All typification[55] as a secondary classifier.
2. Correlations between the actors' responses and researcher-observed hacktivist operations
3. Overall sentiment on misinformation as it exists in their context and the path forward.
4. Tools, Tactics and Procedures mentioned or utilized by the cyber actors interviewed and any commonalities they share with the DISARM framework.

While soliciting interviews, the researcher observed an active counter-operations and investigated the following:

1. Classification of the misinformation diffused.

2. Profile of the misinformation actor and the counter-operators
3. Classification of the counter operation itself
4. Misinformation actor response to the counter-operations

1.6 Assumptions

1. Actors in these communities will be willing to participate but will require the building of trust relationships with eligible survey candidates.
2. Survey participants may prioritize privacy such that demographic information or first party attribution may prove difficult.
3. Survey candidates are willing, good-faith participants who will not require additional compensation.

1.7 Limitations

1. The researcher's presence in these communities was suspect to some candidates; the researcher weighed the effort required to earn their trust against its potential to affect the objectivity of the research and their own personal safety.
2. The volume of responses was limited to the researcher's ability to gain the trust of interested candidates.
3. Non-English-speaking cyber actors did not engage the researcher.
4. Several survey participants are first or second-degree connections, dissemination of information and discussion this research may have taken place, affecting other responses.
5. Survey participants used the survey as an opportunity as an opportunity to attack someone they perceive as spreading misinformation or acting contrary to their ideology. Those responses were removed from the data.

1.8 Delimitations

1. The researcher did not study how eligible candidates or survey participants conceptualize misinformation research or interactions with researchers.
2. This work is *not* an endorsement of hacktivism, hacking, misinformation, or any counter-operations.
3. The researcher did not participate in any ongoing operations.
4. The researcher did not analyze the quality or quantity of any of these cyber communities with which they are embedded.
5. The product of this research is not actionable intelligence nor is it an indicator of compromise.
6. This body of work was not collected using any automated or generative means.
7. The survey with which candidates were presented was furnished as plain-text or spoken during recorded interviews.

Chapter 2: Review of Literature

2.1 The culture of hacking

2.1.1 Hacker Origins

Steven Levy's portrayal of the hacker culture in his 1984 book *Hackers* remains the most influential reference to the public's general view of hackers [28, 20]. Recasting them as Robin Hood-style activists committed to a democratic vision of the Internet [37], Levy asserts that the hacker ethos embodies several sacrosanct postulates to the public good, notably that (i) *all information should be free*, and (ii) *authority should be mistrusted and decentralization promoted* [28]. In the same year "2600: The Hacker Quarterly" is launched, sharing techniques and stories from legendary hackers of their time. Named for the 2600Hz frequency tone which could be used to gain access to AT&T's Operator mode, this magazine became a formative staple in the burgeoning hacker counterculture movement.

Four years on from Levy's book, in 1988, 2600 magazine is compiling it is fifth volume and legend-to-be Kevin Mitnick is imprisoned for hacking into a private company, copying their proprietary software. This same year Mike Pondsmith publishes the first iteration of his tabletop role-playing game series, ”Cyberpunk” and Ken Goffman's magazine ”High Frontiers” is renamed ”Reality Hackers,” canonizing the relationship between counterculture and hacking. Yet another four years on, in 1992, hackers are now firmly enshrined in pop-culture as cool outsiders of exceptional skill with financial gain and personal privacy on their mind. Phil Robinson's movie ”Sneakers” and Neal Stephenson's ”Snow Crash” portray freelancers who exist on the fringes, armed with technology of their own creation, always one step ahead of their narrative adversary. With all this, the do-it-yourself ethos of 1980's punk counterculture and its dramatic political activism has become a foundational cornerstone of hacker culture.

Later-day Internet hackers shifted the ideological tendency for autonomy in cyberspace towards a vision of the Internet as a popular space for sharing information with an understanding that information can be politicized and weaponized against the neoliberal elites responsible for economic and social disarray [15]. Internet activism as a form of socio-political resistance online [26], enabled a functional selection of issues that no longer necessitated a long preparation [32]. This, in turn, resulted in almost instant convergence and coordination of activities in response to the issues of interest that, over the years, became publicly visible through mass media coverage [21].

2.1.2 Activism Online

Internet activism, expectedly, bifurcated to online campaigns concerned with the protection of the Internet as a relatively unregulated and unowned space (e.g. Electronic Frontiers Foundation, WikiLeaks, Snowden [6, 49, 50]) and online campaigns concerned with the protection of human rights and the environment (e.g. Anonymous, the Occupy movement, Arab Spring, Pirate Party [35, 27]). The former activism – or *hacktivism* often is anonymous, performed in secret by small groups of actors, and operates with a kind of impunity that the Internet technologies seem to afford so far [51]. The later activism – or

hashtag activism – usually is public, openly using third party systems on the open Internet for political mobilization, operates primarily on social media, seeking to mobilize participants into the streets, and is subject to the dangers of crowd violence, harassment, and arbitrary arrest [39].

The hashtag activism historically utilized various Internet technologies like petition websites (e.g., MoveOn.org or Change.org) or e-mail communication (e.g. Political Action mailing lists) [5], but the advent of social media sites like Twitter, Facebook, and YouTube truly accelerated the self-organization and participation in the sociopolitical struggle (e.g. the #BlackLivesMatter and #SchoolStrike4Climate movements [13]). While the essential dependence on social media is apparent, both in a historical context and for the future of the hashtag activism [25], the relationship between hackers, activism and social media is a bit more complicated.

2.1.3 Hacktivism

Hacker activists (hereafter ”Hacktivists”), in contrast, exploited various Internet technologies to deface websites [38], to breach systems and “leak” or “dox” private documents [49, 52], and storm systems with traffic to cause a Denial-of-Service (DOS) [34] in the name of ideological grievance. Hacktivists’ foray in social media mirrors these actions as campaigns were undertaken for hijacking/defacement of social media accounts (e.g., Anonymous #OpKKK campaign [54]), doxing individuals on Twitter (e.g., the students at Covington High School [29]), and DoS Twitter topics (e.g., #IranTalks campaign [36]). But hacktivists also exploited the social media affordances for content amplification (e.g., StayWokeBot [14, 41]), early instances of trolling (e.g., Rickrolls [40]), and sharing memes (e.g., Lol Cats on 4chan [6]).

Despite the intuitive versatility of social media for such subversive operations, hacktivism became inactive on the mainstream platforms following high-profile run-ins with the legal authorities of the leading hacktivists [24, 53]. As the operations of Anonymous and their ilk begin to cool in 2016, state-sponsored actors and conspiracy theorists begin using the hacktivist playbook on social media for actions aimed not at elites but the entire social order [11]. Malicious actors use networks of semi-automated

accounts are to exploit engagement algorithms, forcing platforms to signal boost their message and drive recruitment to their chosen cause. Primarily focused on political trolling and sharing of memes, the colonization of social media hacktivism by these actors can be observed during the Brexit campaign in the UK [8] and the 2016 elections in the US [2]. The crucial difference in these instances was that the amplified memes and trolling were not pranks but damaging fake news, emotionally charged memes, and conspiracy theories.

These actors eschew the hacker ideal of a democratic, open internet, instead dividing users and hackers alike in to do battle in cultural warfare. [48]. In response to such a large-scale disruption on the social media turf, one would have expected that the hacktivists will retaliate and confront, expose, or counter hack the state-sponsored "trolls" [59]. Misinformation, back to the Levy's depiction of hacker's ethics [28], runs counter the (i) *all information should be free* postulate because it undermines the basic utility of information as a public good (i.e., truth and facts do not dwindle in supply as more people "consume" them and truth and facts are available to all people in a society) [10].

Misinformation also runs counter the (ii) *authority should be mistrusted, and decentralization promoted* postulate because it is promulgated by a state-sponsored "shadow authority," as evidence confirms in the aftermath of the Brexit and the 2016 US elections [31, 56, 22]. Surprisingly, the hacktivists never struck back [3], though they clearly pose the capabilities to do so, as witnessed in the Anonymous' #OpISIS campaign, for instance, where the collective flagged about 101,000 Twitter accounts attributed to the Islamic-State [23]. The absence of response to misinformation on social media by the hacktivist community seemed quite perplexing and worthy of in-depth inquiry with active "hackers" that still operate in the spirit of the Levy's code of ethics [28].

Where a hacker might exploit the weakness of a certain system's memory handling to gain persistence, so too do these actors exploit the public's short memory to gain persistence in our worldview. The perpetual motion of the 24-hour news cycle is a brute-force attack on the public's psyche and the individual, before social media, had extraordinarily little recourse to engage the deluge of information. The individual consumer has no mitigation other than to willfully choose not to consume. Where a

journalist would inform the public of significant information, the Influencer uses that information to inspire the public to act on their behalf. Where the news media and the Influencers have platforms on social media, their use of it differentiates strongly. Counter-culture Influencers use tribal communications, laced with vernacular, posting familiar appeals to the intended consumer's worldview using the default systems of social media.

News-media and Political parties are beholden to governance and thus restricted to specific modes and content, where violation is supposed to result in repercussions. The stakes are much lower for the Influencer on social media, being banned from the platform does not prevent one from creating new accounts and re-connecting with your consumer base. Where we might create mitigation to prevent the unauthorized access to private systems, to prevent hackers implies one does not have authorization to access the system in which information lies, the Influencer not only has access but is also using approved channels. Both the hacker, the activist and influencer have goals with which they execute on throughout various systems to realize ideological or systematic change by exposing information and casting daylight on difficult concepts for the general public's consumption.

2.2 Misinformation

2.2.1 Misinformation's historic parallels

Modern media uses influencer as an ephemeral term for social media professionals, or people who use their reach to influence the consumption of some product or ideology. However, this is just a modern iteration on the broadcast personality and some of their tactics are used to this day. Norman G. Baker is one such blueprint for the modern misinformation actor. In 1930 he used his radio station KTNT to advertise fraudulent medical treatments available only at the hospital he owned, while sowing distrust in the local medical establishment of Muscatine, Iowa. Baker asserted that trained professionals were exclusively motivated by money and not the well-being of the patients.

Through the course of the COVID 19 pandemic influencers have used their vast reach to espouse the use of untested treatment methods as alternatives to vaccination.

Misinformation actors attacked the pharmaceutical industry and state civic health bodies as exclusively financial, discrediting involvement their involvement as a sinister often offering links to storefronts where their favorite cure is sold. Famous conspiracy theorist and entertainer Alex Jones appends all his tales on why the listening consumers should purchase his colloidal silver cures or disaster recovery kits.

Baker was not alone in his abuse of early broadcasting to generate fear in the consumer audience while selling a product and accumulating political and financial capital. In 1920, John Brinkley was lauded by local press for the success of his goat-gland transplant in a pregnant patient. Soon after 1923, Baker took to Kansas KFKB radio for four hours a day promoting his goat gland treatments, appealing to listeners emotional response and ego. Responding to public outcry, Brinkley runs a populist campaign for the Kansas gubernatorial election of 1930. Losing that, Brinkley sells KFKB and moves to the Mexican border, establishing a new, high power radio station with he which broadcasts a new political campaign and more miracle cures.

The geographic location of John Brinkley's radio station in Mexico allowed him to circumvent a number of communications controls recently placed on radio broadcasts in the United States. Using a telephone to call and then broadcast his message over what was, at the time, the most powerful radio antenna on the planet, there was little that could be done to stop Brinkley from spreading his dangerous and fraudulent information. That is, until 1932 when Congress passed a law named for him, the Brinkley Act, banning the practice. Brinkley eventually lost his broadcast license in Mexico, under pressure from the United States, but his fraudulent activity did not stop until he was finally prosecuted for them. External pressures such as the application of the justice system, have historical precedent as forward-looking misinformation countermeasures.

Similarly, banned celebrities or influencers ousted from their dominant social media platforms of choice publicly went on to create new platforms with more lax moderation; where the very narratives that had them removed from the previous platforms were promoted as primary content. After twitter

suspended, then banned former President Trump, he founded an alternative platform called Truth Social, broadcasting his political messaging to a captive fan base. While Trump could no longer reach his base, screen captures of his Truth Social posts are readily shared on other, platforms evading punishment.

2.2.2 The comments section.

In 1998 The Rocky Mountain News opened the first comments section on the web to the public. From the Comments Section sprung forth a pervasive tradition of opining on the truth and voicing mistrust of the publication. It was not long before Myspace (2003) and Facebook (2004) were founded, and the comments section now existed under every activity a consumer admits to. Transposing the comments section from beneath the article to the first thing you see, large audiences, formerly the domain of mass-media, are free to all who can cultivate one. Historical archetypes fond of microphones such as propagandists, street evangelists, media pundits and celebrities capitalize on the newfound access to the public. These archetypes persist as the Information Activist begins to develop alongside Hacktivists and the Post-Social Media web.

There exposes may have previously been the domain of the plucky traditional journalist, social media is now home to the "citizen journalist" and hobbyist whistle-blower. These reductive archetypes outline a trend in the contextualization and dissemination of publicly available or, at least, not explicitly illegal information as a means of action on their ideological adversary. Influencers and journalists differentiate in an important way: where a journalist should aim for objective reporting of events, the Influencer uses cherrypicked details of events and presents them, colored by their ideology or agenda as objective reporting. The volume with which information was distributed to downplaying the 2020 COVID19 pandemic underscores the willingness of the general populace to traffic in falsehoods and potentially harmful concepts. In the comments of news articles and a burgeoning YouTube, a proud tradition of ad-hominem attacks and accusatory language was formed.

2.2.3 You can be anything on the Internet.

Denizens of the Internet might gleefully dismiss the web as a poor source of credibility, joking that "you can be anything you want on the Internet." However, there is a sense of ownership over what is posted on one's favorite platform. This detachment caused by social pressures to be honest and the ease one may lie or deceive their fellow consumers created a dangerous atmosphere of willful mistruths for the sake of willingly aggravating one's ideological adversary. Through the duration of the COVID-19 pandemic, influencers and the public have openly traded in deceit to create narratives that affirm their worldview. In misinformation we are presented with the convergence of data, community, and information; all three concepts are native to the hacker counterculture's foundation.

Where an early hacker might tell you that information yearns to be free, they are not commenting that the value of information decreases if it is found to be fraudulent or incomplete. The success of spreading misinformation on social media shows us an inverse relationship: value in this system is a high engagement rate and, if information on social media creates engagement with consumers, its value is greater than less evocative truths. Where information security first principles can be applied to traditional telecommunications and networked systems, that is only if the networks and systems are bound by the rules or instructions they were given. Network communications trend towards efficiency taking the shortest path to their destination, on social media, the destination is the user's engagement.

Misinformation's reliance on the emotional response and the consumers inherent biases are like the playbook of a modern phishing kit. A phishing operation seeks to target as many consumers as possible, evoke emotional response that will lead to further engagement and, compromise the consumer & their network to pursue objectives. Misinformation on social media is a distributed attack on the veracity of a given topic in the ongoing public discourse, aimed at as many consumers as possible within a given social network. Where a cyber threat actor might deploy a worm to propagate to network endpoints, so too do the actors willingly participating deploy memetic or viral content as they diffuse misinformation into an online social network.

The DISARM Red framework acknowledges the power of state-level actors, charlatans, or influencers as powerful vectors of misinformation diffusion. Between 2022 and 2023, the resurgence of chauvinist life coaches on platforms such as TikTok, Rumble and YouTube applied advice on successful living using absolutist language and emotionally charged rhetoric. Actors like Andrew Tate used short clips diffused throughout TikTok and Instagram, to engage young men seeking financial and romantic success but, only if they subscribe to the Tate philosophy, both monetarily and ideologically. Establishing a referral network, and a "Hustlers University" in a series of Discord servers, Tate and his brother were able to generate significant revenue by extracting $50 a month, for invitations to a given server where influencers would teach them drop shipping, cryptocurrency trading, and other non-traditional forms of revenue generation.

The Tate brothers have no training in psychotherapy or day trading, instead asserting their material success is proof of their credibility as gurus, laundering their image before young impressionable consumers. This false expertise portrayed by Tate is never backed up with any evidence other than content highlighting his material wealth. We now know Andrew Tate was a prolific sex trafficker in Romania; but his indoctrinated community asserts that he had been framed simply because he spoke the truth. These confirmation biases are infectious, and the Andrew Tate example is easy to replicate.

2.3 Counter-culture Actors

Where the transit of information is done using data, over Ethernet, fiber and microwave, there is an understood abstraction between that information and the data that carries it. To many, "Online" is not "real life" and thus, the consequences for misbehavior in social spaces are not as great as they might be in a physical social setting. The hacker and influencer reject this, online *is* real life; community, subsistence, all extant online and, easy to source for most hackers. There are "real life" consequences for their actions; now, too, there are consequences for social behaviors that are not explicitly a breach of a digital system. If the internet is not "real life" then why are attacks on one's character so persistent and damaging? Look to

the work of Katelyn Bowden, Holly Jacobs, and their contemporaries as proof: forty-six states have laws prohibiting Revenge Porn, criminalizing the act of putting ”real life” online.

Attitudes have begun to change as to the importance and impact hackers can have on quality of our infrastructure. In 2022, the US Justice Department announced that it will not be bringing charges under federal computer crime laws against security researchers acting in good faith. This removal of the taboo, namely the act of hacking being a property-based crime, is a tacit admission by the Department of Justice that outsiders create value, if their activities serve the civic good of our infrastructure. The Cybersecurity and Infrastructure Security Agency (CISA) further legitimizes cyber-actors as important fixtures in civil society; breaking out the roles of cyber talent from the traditional government pay and hiring structures.

Furthermore, pop culture and online counterculture accept online friends as real-life friends. Social media encourages you to live your life in the open if it serves the interest of their advertisers. So, too do social media advertisers acknowledge this concept, both on their bottom line and with the public good will towards their brand. Notably, YouTube is known to arbitrarily demonetize community content should advertisers find the topic, or at least the keywords, objectionable or damaging to their public image. In this way, YouTube and those advertisers acknowledge that social media content is just as valuable as traditional broadcast or physical advertisement.

Conversely, bad faith actors, such as Tate, who are unskilled in OSINT methodology assume the layers of abstraction presented by social media protect them from their immediate audience discovering their truths. Those who do have the skills, such as journalists, hacktivists, and intelligence hobbyists, they are more than willing to destroy those layers of abstraction by holding these parties accountable with freely available and credible information. An actor like Amnon Bundy, who led the anti-COVID-19 restrictions trucking and shipping group “the Freedom Convoy,” asserted that their success and righteousness was evidence of their movement’s credibility. Aubrey Cottle responded by exposing their source of income by leaking the donor list of their crowdfunding campaign on the Christian-centric alt-platform GiveSendGo, revealing their donors to be mostly uninvolved with the actual convoy, Canada, or independent trucking.

No longer is the use of any particular social platform a barrier between charlatans and accountability. Between 2021 and 2023, self-styled mobile forensics experts Jonathan Scott a.k.a. JonathanData1, abused the academic social media platform ResearchGate to publish counter-narratives to Amnesty International reporting on the state-level spying perpetuated against opposition movements within the Catalan and Moroccan citizenry. Asserting that he was a published author, Jonathan demanded respect and refused to engage with criticism, unless they were similarly credible in the field of mobile forensics. Content to levy ad hominem attacks against his detractors on Twitter.

Using open-source intelligence and outsider expertise, journalist Runa Sandvik, Android Security engineer Lukasz Siewierski and, a group of activists known as TeamTaco have been working discredit and reduce the engagement Scott has cultivated. Their aim is to limit Scott's ability to diffuse misinformation into mainstream media and limit his attacks on the information security community. This external and community-led effort is ongoing but continues to attract the attention of other similarly inclined members of their respective communities. However, the platforms where said groups convene and operate have yet to condone, condemn or enable the misinformation counter operations. Rather than act, these platforms tacitly endorse injurious behavior, as long as it the diffusing actor's wider behavior patterns are observant of their respective terms of service. Where these disinformation actors are met with public outcry and external pressures, such as regulation or loss of revenue, their ongoing efforts become frustrated, lose momentum, and become starved for engagement.

Chapter 3: Methodology

To ensure validity to the task of conceptualizing misinformation, the researcher introduced the participants in the main study to the generalized definition of misinformation on social media proposed by Wu et al. [55]. The cyber actors in our sample were invited to speak about their profiles, activity, agendas online, and their take on misinformation's presence on social media. The qualitative responses were coded and categorized in respect: a) antecedents to misinformation; b) mental models of

misinformation; c) countering misinformation through leaking, doxing, and deplatforming; d) counter misinformation operations ("ops"); e) counter-misinformation tactics via DISARM Blue. f) misinformation literacy; and g) misinformation hacktivism.

3.1 Sample

Subjects were identified by embedding the researcher in active hacker & information security social spaces on Twitter, Signal and Discord. Where activity is judged by frequency of posts to relative size of the active community, the researcher observed the Anonymous, VX-Underground, Cult of the Dead Cow and Telecomix communities. Participants were largely active in U.S. time zones with explicit, first or second-degree connections to the following online collectives.

3.1.1 Sample Communities

1. **Anonymous -** A collective known for its decentralized approach to combating misinformation and waging cognitive battles online.[6] While anonymity is portrayed as a hallmark, these communities formed around strong personalities who openly and willingly associate with anonymous online, first party association with Aubrey Cottle (founder 420chan) or Frederick Brennan (founder 8chan) is a clear indicator of interest in cyber-activism or security research.

2. **Cult of the Dead Cow** aka "cDc" – A collective known for embracing counterculture and hacking from a community-led perspective, which embraces unique personas over anonymity.[33]

3. **Telecomix** - An affiliate of Anonymous, Telecomix is a collective of hacktivists who actively engage in psychological operations to push leftist propaganda.

4. **DefCon Attendees** - seen as the celebration of hacker counterculture this conference draws attendees from all over the world and serves as a nexus for crosspollination between communities and collectives. Attendees vary between law enforcement, cyber-criminal and otherwise interested party. Attendance or association with attendants is a reliable indicator of a user's interest in cyber activities.

5. **VX-Underground** - Virus Exchange Underground or VXUG is a not-for-profit organization that collects and hosts one of the world's largest collections of malware samples, physical and information security research. It is community on Discord and Twitter are an international mix of English & Russian speaking information security professionals, malware researchers and erstwhile cybercriminals.

6. **NAFO** – the North Atlantic Fellas Organization or #NAFO is a social media movement dedicated to countering Russian disinformation about the 2022 invasion of Ukraine.[42] There is a considerable amount overlap between Anonymous, Telecomix, DefCon Attendees and the OSINT community at large. Willingly identifying with this movement was a strong indication of active involvement in misinformation counter-operations.

3.1.2 Survey Candidacy

Candidacy for interview was established if a community member displayed three or more of the following criteria:

1. Subject was active in the social space daily.
2. Subject engaged in conversations about misinformation openly.
3. Subject openly identified as interested or actively engaging in hacking, OSINT operations, cybersecurity, malware, security research or free-internet activism.
4. Subject frequently engaged from an ideological or experiential viewpoint when discussing misinformation or cyber-activities online.
5. Subject displayed affinity for community-led or academic research.
6. Subject shared tools, tactics or news items related to information collection or information security at least once a week.
7. Subject observably maintained first or second-degree association with Anonymous, VX-Underground, Cult of the Dead Cow, NAFO or Telecomix-affiliated actors and

social spaces.

3.2 Data Collection

Modes of collection included video call, Twitter & Discord direct messaging, and e-mail. The researcher posed the following questions to each survey participant, and they were encouraged to answer in as much detail as they were comfortable giving.

1. How do you describe your niche, role, activity, or agenda you have online?
2. What brought you to hacking, OSINT, cyber-threat intelligence, and any operations you have taken so far?
3. Have you faced any obstacles, challenges, repercussions because of your activity?
4. Has the obstacles, challenges, repercussions affected your commitment, motivation, and vision of your actions and in what way?
5. What is your take on the increased misinformation proliferation online?
6. Have you ever engaged or considered engaging in utilizing your actions in exposing disinformation campaigns? What was the disinformation about, in what capacity you participated, and what were the outcomes you were attempting to achieve?
7. We would like you to present a hypothetical case of operational OSINT to uncover a COVID-19 disinformation campaign. What is your take on doing such an operation in the first place? Would you have done it or participate in some capacity?
8. What do you think the tools, tactics, and procedures undertaken in a hypothetical *misinformation hacktivism* operation might entail?
9. What in your opinion, is the way to continue evolving this work and in what shape and form?
10. Is there anything else that you would like to add or say that is relevant to the questions we have asked so far?
11. If you would like to share some demographic information, please do - we do not require it, but it will help us better contextualize your effort and story.

3.3 Data Analysis

Two independent researchers analyzed the survey transcriptions to achieve a strong level of inter-coder agreement (Cohen's $\kappa = .82$). The researcher then utilized a thematic analysis methodology to identify the themes and sub-themes most saliently emergent from the responses collected. The themes are summarized to describe the conceptualization, response, and evolution of misinformation in the view of the contemporary cyber-actors surveyed. To report the results, the researcher utilized as much as possible verbatim quotation of participants' answers, placed in "quotations" and with a reference to the participant as either **PX**, where **P** denotes **participant** and **X** denotes the **number** of the participant in the sample.

Chapter 4: Presentation of Data

What follows is a thematic, qualitative analysis of 25 Participants responses to the Eleven question ethnographic survey. Three participants chose to respond in video-call interviews and the remainder responded in text-based short-answer surveys, via e-mail, Discord and Twitter direct messages.

4.1 Survey Results

4.1.1 Question 1

How do you describe your niche, role, activity, or agenda you have online?

FIGURE 1. Q1 TOTALS

CYBERSECURITY	10
ADVOCACY	10
DFIR	2
THREAT INTELLIGENCE	3
INFLUENCER	8
HACK/TIVIST	4
HUMAN INTELLIGENCE	4
EDUCATION	4
RESEARCHER	6
DEVELOPER	4
COMMUNITY LEADERSHIP	3

Of the twenty-five (25) surveyed, all but two used multiple terms to classify what their presence online means to them. Of the more colorful definitions, P4 classified their current role as a "3rd shift Firefighter", asserting that their cybersecurity work is not their primary function online; advocating for Socialism in their sphere of influence was also important to their identity. The theme of advocacy occurs ten (10) times in total, ranging from freedom of information (4 occurrences) to privacy (2 occurrences), to

conflict and extremism watchdogs (1 occurrence, each). Nine respondents identified as Cybersecurity professionals, many of whom added other themes to their roles such as Advocacy (4 occurrences) and Digital Forensics & Incident Response (2 occurrences). Six (6) respondents classified their activities as Influencer, notably paired with Researcher and Educator twice, independently.

The researcher classification occurred five (5) times, grouped with Influencer (2 occurrences), one respondent categorizing their multitude of talents, P17 “combination of evangelist, trainer and ... propagandist”. Four (4) respondents classified their activities as Human Intelligence gathering (HUMINT), none of those four sharing similar classifications, adding Cybersecurity, Social Engineering, Community Defense and Research to their roles, independent of each other. Another four (4) respondents categorized their activities as Threat Intelligence be it for the Ukraine conflict (1 occurrence), Threat groups (2 occurrences) or malware; 3 of these CTI specialists also classified their operations as Research. Notably, twenty-three (23) respondents classified their activities as two or more of the preceding categories, sentiments like P24’s “Jack of All Trades” are common among respondents. Of the three who used singular definitions, P8’s “Hacktivism” and P12’s “Hacking” stands out; the simple self-classification as a “hacker” is rare, in our context, and our respondents have complex definitions of their activity, online.

4.1.2 Question 2

What brought you to hacking, OSINT, cyber-threat intelligence and any operations you have participated in, so far?

FIGURE 2. Q2 TOTALS

YOUNG START	8
TEEN START	6
HIGHER-ED	1
CAREER CHANGE	3
SELF-LED	6

Here, respondents became more verbose, adding complexity to their personal contexts as cyber-actors. While no two stories were explicitly the same, there are a number of commonalities worth noting. First, we observe the “*when*”: Many took the opportunity to offer what age, phase or style of learning began their journey to their current cyber-role, which the researcher classified as: Young Start (between the ages of 9 and 13), Teen Start (14 to 19), Higher Education (post-secondary) and Career-

Changers. Of those roles, eight (8) identified as Young Starters, six (6) as Teen Starters, two (2) as Higher Ed. learners, three (3) identified as career-changers and six (6) respondents in this group classified some part of their journey as Self-Led Learning.

What brought the respondents to their cyber-roles is as varied and overlapping as the preceding characterizations. Video Games were mentioned 7 times, either in the pursuit of cheating, piracy, playing online or, getting the most performance out of their PCs. Seven (7) mentions of Community Leadership were made, referring to their prominence in a specific community being either the reason they honed their skills or the result of their exceptional talent. Respondent P3 notes that their community leadership was borne of their social engineering work using "group manipulation techniques to co-op corrupted small communities to allow groups to evolve freely."

The *Why* for our respondents is a bit more varied but, terms used in the first question reoccur. Advocacy receives 4 distinct mentions, with similar sub-classifications, Victim Advocacy occurring twice and eliciting strong responses. P6F expands on their OSINT skills here explaining, "girls have to creep on new guys they go out with 'cause... we've heard stories or experienced first-hand a guy being an absolute dirt bag in real life and we could have easily worked it out of his social media," using this example they go on to note that "sexism is a vulnerability." Previously, only one respondent identified their role as "hacktivism" but here, three respondents identify it as part of their origins. Respondent P7 expands on how hacktivism drove their developing HUMINT and Social Engineering skills, divulging traumatic experience, that "watching similar things happen to vulnerable people, children and other people, made me want to utilize what I knew to make sure no one would have to feel powerless once they'd been targeted."

Respondent P8 asserts a wider scope, becoming agitated as they remark on "human rights abuses" observed "in Russia and China" as the impetus for developing their "hacking" skills. Other respondents remark on puzzles and problem-solving as motivation. Respondent P18 remarks that it is the logos for their role online remarking, "the whole reason I got into computers is because from a very young age, I liked finding solutions to problems." So too does P18M remarking that their

research is because of their “interest in understanding how things work.” P21 explains a that their journey “began as a puzzle solving [Capture-the-Flag]-type situation whilst in college.” For all this diversity in origin, P16M astutely underscores an important commonality in the respondent’s personalities, “Many people, in this field all have something in common and it’s curiosity.”

4.1.3 Question 3

Have you faced any obstacles, challenges, repercussions because of your activity?

FIGURE 3. Q3 TOTALS

YES	15
NO	8

This question is closer to presenting a unifying picture of the respondents, 15 of whom responded in the affirmative, stating those repercussions were detrimental to their well-being, online and in the material realm. Three (3) respondents admit to being contacted by Law Enforcement, all asserting that did not dissuade them from further activities; heightening their operational security awareness or forcing them to ground for some period.

Law Enforcement was not the only material threat, P15 explains that they experienced “the occasional threat from a fascist or racist group” but were undeterred, elaborating that “unmasking them and exposing who they are, which affords them real-world consequences” was reward enough. Acknowledging the risk, they take is because of their opposition to predatory actors, on behalf of their victims P17 explains “...some of these people have been targeted and I feel like even if I do get exposed, it wouldn’t be the end of the world.”

Some adverse effects are experiential, P25 remarks that being a technology influencer “has negatively affected my real-life relationships.” Where P21 feels “siloed into specific areas” because of their activity, P13 admits to experiencing “imposter syndrome”. Further, P5 admits they “have had to change identities online many times,” which is troubling for their credibility since “history is how people are vetted in anonymous communities,” experiencing distrust in their social circle due to their “relatively short one.” P24 explains that they have to play down their experience, “I usually just mention that I work in IT to avoid being adversely labeled.”

Of 7 respondents who offered a clear “No,” only 2 elaborate further. Respondent P16 explains that is because they are privacy-forward, operating “underground and discrete.” Expanding on that concept, P18 states that the “closest thing to an obstacle I have is maintaining operational security associated with things my job pays me for, not making dumb mistakes *in situ*, and keeping a functional sleep schedule.”

4.1.4 Question 4

Have the obstacles, challenges, repercussions affected your commitment, motivation, & vision of your actions and, in what way?

FIGURE 4. Q4 TOTALS

INSPIRED	8
ANGRY	2
FRUSTRATED	3
DETRIMENTAL	3
RESOLUTE	6
“NO”	3
AFFIRMING	2

Sentiment further coagulates here, eight (8) respondents characterize their past challenging experiences as inspiring. Respondent P23 distills their experience neatly, remarking, “It just makes me go harder,” and P6 agrees, with a touch of anger, “It's made me louder. More committed." Similar messaging comes from P15, who remarks that obstacles “tend to make me push harder for what I believe in.” The intentional attitude is pervasive, P3 remarks that when “no one seems to care in the wider world you realize at some point it’s time to get busy.” In furtherance of their own mission, P13’s inspiration is their opponents, “they have motivated me to be better at my job.”

Six (6) respondents use language indicating resolve, neither particularly inspired nor detracted, content to keep pace with their personal mission. P7 paints a picture of growth, “I don’t think punishment or other people’s revenge will ever change what I believe in, but it may change the way I do things." P9 remarks that it is not willpower but absence of risk, adding “maybe if the police were called, I would be more paranoid to this day.” Further, P24 wishes for longevity, “I just want to take it easy and grow roots which is exactly what I've done." A less romantic resolve is offered by P18, who comments on the frustrations created by honing the skills of their community, “I remind

myself of the many advances in the security industry that have trickled down to protecting consumers."

Others have less positive sentiment, P8 remarks on the loss of a comrade's life. Some find it difficult to connect with their achievements, P22 decides to "detach from activism" and P21 feels the "privilege of [others] ... makes it hard to compete so, I ended up just... settling for [less]." Challenges have pushed P14 away from their peers, noting that "Knowing who you can trust is key" to their longevity. Finally, P4 feels frustrated with the opposition they've faced noting that, "hard to care about trying to make things better when absolutely no one gives a shit about what you're trying to do

4.1.5 Question 5

What is your take on the increased d/misinformation proliferation online?

Responses to this question are the most varied, so far. Eleven (11) respondents view misinformation as external to themselves, offering a median of four (4) additional classifiers, or ~4.18 on average. Three (3) of whom remark that misinformation is a threat to our collective cognitive health or, as P5 puts it, "Information has been manipulated in a way that is actively destroying people's brains." Four (4) in this group believe misinformation to be external propaganda, as P7 observes, "Whether it's 5G giving everyone cancer, there being nanobots in vaccinations... [misinformation] ends up making people anxious rather than achieving anything productive." Two (2) in this group see misinformation as inevitable, P23 comments that proliferation is "To be expected when cowards are able to hide behind the false vale of anonymity."

FIGURE 5. Q5 TOTALS
NOTE: DUE TO HIGH LEVELS OF VARIANCE, ONLY CLASSIFICATIONS WITH THREE (3) OR GREATER RESPONSES ARE DISPLAYED.

EXTERNAL TO RESPONDENT	11
RESPONDENT OPPOSES	8
EXTERNAL PROPAGANDA	7
BIASED	5
LACK/FAILING OF EDUCATION	4
CONSUMED BY RESPONDENT	3
RESPONDENT IS AVERSE/AVOIDANT	3
POLITICAL (COUNTER) ARGUMENTATION	3
THREAT TO COGNITIVE HEALTH	3
PERPETUATED DUE TO USER BIASES	3

P16 comments on misinformation's historic precedents remarking that "the Internet is just the latest means of dissemination." Eight (8) respondents assert they are opponents of misinformation offering a median of two (2) additional classifiers and, ~2.8 on average. Additional to modeling misinformation,

respondents often give their opinion on who they feel is responsible for the proliferation of misinformation, beyond just actors who perpetrate it. Four respondents (4) indicate failures of education, either that of the consumer or the system itself, are to blame for the proliferation of misinformation.

However, three (3) respondents express an aversion or ambivalence to the phenomenon, P8 remarks on behalf of their collective, “We don't have an opinion on it and we're not doing anything about it... leave it to some others who may be more skilled." Where P9 takes an ideological stance, “Never restrict the flow of information, not even d/misinformation.” P21 expresses ambivalence stating that misinformation “has always been here,” declining to elaborate.

4.1.6 Question 6

Have you ever engaged or considered engaging in utilizing your actions in exposing disinformation campaigns? What was the disinformation about, in what capacity you participated, and what were the outcomes you were attempting to achieve?

In this series of responses, there is a triarchy; four (4) respondents were active, but are not currently, ten (10) are currently active and eleven (11) do not or would not participate in counter-disinformation operations. Active participation skews male-identifying, ten (10) male respondents are or were active recently. Nine (9) respondents make mention of operational tactics that directly map to the DISARM BLUE and RED framework. Three (3) respondents indicate RED (misinformation perpetrator) tactics as part of their counter-offensive toolkit. Respondent P6 & P24 admit to the use the use of T0048.04, Doxing, as weaponized against predators, P24 explains their sentiment, “Some of these clowns have it coming to them." Similar aggressive sentiment is expressed by P14 who responds to the question plainly, “Exposing crypto scammers is better than sex."

FIGURE 6. Q6 TOTALS

ACTIVE	10
INACTIVE	4
DO NOT/ WOULD NOT	11

Other counter-operational respondents offer further insight into operations they are familiar with or have participated in. P16 divulges one operation they are proud of, “OpJane, also known as Operation Jane, was in response to Texas Government making abortions inaccessible as well as their actions

intended to encroach on the individual freedom of women." P22 is expresses a more local, personal angle, “I address misinformation... in the communities I help run and support." P23 zooms out further simply stating that they “usually get pretty involved with things." Even more external, P17 speaks on the tactics of AM radio misinformation, “Sometimes the narrative can be intentionally confusing... the goal is to try to get the Listener to feel stupid.” P7 explains their work is goal-based on behalf of the victims of misinformation, detailing their experiences with T0105.001 (using memes in furtherance of a message), C00133 (deplatforming), and C00130 (mentorship); they summarize with “My desired outcome is always for people to build opinions based on facts and with an understanding of all perspectives."

The group that does not or will not participate is more evenly distributed: five (5) masculine-identifying, five (5) feminine-identifying and one (1) non-binary identifying. Some respondents’ express cynicism: P4 explains counter-operations are “a waste of time... The views of people online at this point are, in my opinion, set in stone.” P10 feels similarly, “you’re not accomplishing anything by trying to silence opposing views.” Asserting their ideological stance once more P9 states that platforms should “let the information flow free, even artificial information.” P8 takes a more pragmatic stance, re-affirming their previous statements “Best leave it to some others who may be more skilled.” P11 feels expresses frustration in the third person asserting that more people don’t act on misinformation because “you have to have a massive amount of influence... for [operations] to be effective.” Finally, only one (1) from this group expresses desire to act, in spite of their current operational status, unsure of their own capacity, stating “if I had the opportunity and I felt like I could take it on, I would."

[This space left intentionally blank for formatting purposes]

4.1.7 Question 7

We would like you to present a hypothetical case of operational OSINT used to uncover a COVID-19 disinformation campaign. What is your take on doing such an operation in the first place? Would you have done it or participated in some capacity?

FIGURE 7. Q7 TOTALS

AFFIRMATIVE	14
NEGATIVE	7
INDEFINITE	2
NO RESPONSE	2

Respondents react to this question in the affirmative, with a negative sentiment, indefinite responses or, in the cases of P6 & P8, refusal to comment. Affirmative responses are in the majority, totaling fourteen (14); again, skewing masculine with ten (10) male-identifying respondents and four (4) female-identifying. Of the seven (7) negative responses are again more evenly distributed across gender-identity: four (4) masculine-identifying and three (3) feminine identifying.

Some affirmative replies are direct: P5 “would applaud such an operation”, P13 says “I think people stepping in, to help is going to be helpful” and, P15 believes “the fight to expose and stamp out disinformation is central to our society's future." Others offer tactical insight: P17 discusses their use of C00133 (deplatforming) & T0048 (harassment) in their operations and remarks on what they perceive to be endemic issues on social media, “… we gotta stop treating disinformation as freedom of speech. It's one thing to think you can say with what you want but, that shouldn't shelter you from the consequences.” P7 reiterates the importance of C00130 (mentorship), “exposing of misinformation has to be done in an educational and non-confrontational format.”

Some are more willing to dig into tactics. P16 remarks on the efficacy exploiting of technical systems (T0081) of a platform to counter misinformation. P23 briefly details and operation where they employed C00029 and C00154 (ask media not to report false info), remarking that they “would do whatever is needed.” Others are pragmatic, P18 offers “I am happy to help where my talents lie, but the feeling of unclear success criteria or demonstrable outcomes from success makes it hard to feel intensely motivated.” P24 remarks that misinformation actors’ motives must be exposed (C00115), invoking a code of conduct in the first-person, “I would ensure that I wasn't doing anything that could lead to... harm.”

Finally, P25, a tech influencer, approves of this scenario while observing the personal risk, "I wish I could do more but as stated [in the previous response], partaking in anymore online activity would negatively affect my real-life relationships."

The significance of personal risk is not lost on P9, who restates their personal philosophy, "I just want to be left alone, and leave other people alone." P11 also opines on the risk present in these operations, from the third person, "many people will likely become bored and give up, or realize their 'reputation' in the corporate world matters much more than justice and truth." P4 acknowledges the threat to our cognitive health but feels operations are futile, remarking on recent history: "We had a dangerous charlatan as the head of state for 4 years and something like half of the United States still believes that his reelection was stolen from him."

4.1.8 Question 8

What do you think the tools, tactics, and procedures undertaken in this operation might entail?

FIGURE 8. Q8 TOTALS

PRACTICAL	13
TECHNICAL	6

Here, nineteen (19) respondents offer a blend of practical and technical responses, remarking on specific tools & technologies or how one might leverage them during a counter-misinformation operation. The remaining six (6) respondents either do not approve of undertaking this operation and abstain or don't feel comfortable divulging, as P16 puts it, "I don't have a clear and concise answer to that question." The previous gender-identity biases do not present themselves in the same manner, here; all six (6) female-identifying responses give practical replies to this query.

Respondent P7 offers a comprehensive model of multi-planar preparedness plan for online social networks. They identify C00111 (presenting empathetic oppositional views), C00170 (mobilizing state-level resources), C00019 (providing playbooks to identify divisive actors), C00027 (cultivating civility) and C00017 (repair broken social connections) as critical to civic health online. P7 stresses that fairness and education is the path to success, "[Fighting] misinformation is about ensuring that everyone that has an opinion, or a view is well informed about the facts prior to building that view." P20 agrees, asserting

that “if you have lot of info on a social media feed, it’s truth in the same format.” Conversely, respondent P5 suggests subterfuge, identifying T0100.001 using their own words, “befriending and exposing is a solid tactic on any enemy.”

Also starting with the planning phase, P18 identifies a foundational starting point: the capacity for complex data collection, “You're going to want to gather as much info as you can. Multiplatform collection further complicates things if you're not already spun up for collection on them.” At a more microcosmic level, P17 describes their own processes for defusing a charlatan (C00113), admitting to its informality, “I dig deep, I'll kind of go down these sort of rabbit holes. That's when I'll just unveil as much possible information as I can about [the misinformation actor] and, a lot of times, it's not necessarily a formalized process.” P11 discusses using OSINT to discover content with which to engage a payload and debunk it (C00119). They go on to observe that “we have to be careful and not cross the boundary between reporting for truth and exposing their private information to everyone. We don’t really know what everyone else is going to do beyond us.”

Stressing the need for active counter-messaging to preserve the cognitive health of the target consumers P4 suggests a multi-planar approach, identifying C00105 (buy more advertising than the adversary), C00087 (make more noise than the disinformation): “…throw money at problem with the sum of your resulting intelligence at the internet with bots, billboards, and possibly radio ads if you find your targets listen to something specific." P23 and P24 list names of OSINT technologies they use to assess the vulnerability of a mis-informant’s digital assets, “Zetalytics Passive DNS API, Maltego, GreyNoise, Shodan” and, “Buscador Linux OS distro is generally what I'll use for OSINT ops but, there is plenty to find just by Googling information.” To that end, P12 names similar technologies, punctuating it with offensive tactics, “If all else fails send a [malicious document]... and collect as much information as possible (keystrokes, saved passwords, emails, SSH keys, FTP credentials, browser history, friends, family, etc.)”

4.1.9 Question 9

What, in your opinion, is the way to continue to evolve this work and in what shape and form?

Answers to this question are a bit more nebulous than previous items but, receives twenty-one (21) responses, all the same. Six (6) respondents identify open collaboration as a key strategy for anyone who wishes to engage in this work, long term, as long as operatives "use encryption", P4 notes. P22 notes that there should be "open, central repositories of knowledge" relating to historical misinformation and how to combat it. P7 notes that while development is imperative, "We should aim to make people see that educating people on and correcting misinformation is actually beneficial for everyone." P6 expands on that, "causing a pile on isn't ideal, partly because that makes you just as bad as [misinformation actors]." P17 appeals to empathetic education, "if it's a mistake, we help them correct it. If they keep pushing, then we need to list facts and we need to get it out there."

The presence of capital in social spaces as a root-cause is invoked by P3 and P15, who remarks that social media platforms "have to be put into a tight vise and feel the noose around their necks in a way that will threaten their bottom line." P23 suggests incentivizing counter operations, "Finding a way to monetize [counter-operations] so people can actually pay their bills while doing it." More generally, P5 asserts that those interested in counter-operations should "expose those who spread disinformation. We should look at their motives, their influences, and who they are." P18 believes continuance is more important, "the act of resistance and fighting is more important than winning." However, not all are looking towards the future. P24 feels the threat of misinformation is indominable stating, "I don't believe there is any need for evolving this kind of activism. It's always going to be there, looming in the background and the dark corners of the internet. P9 agrees, the threat is ever-present, "I don't think it will stop."

4.1.10 Question 10

Is there anything else that you would like to add or say that is relevant to the questions we have asked so far?

The responses to this question are anemic, thirteen (13) participants choose not to reply. Two (2) respondents remark on capitalism's intrusion into our social spaces, expanding on social injustices they perceive. P3 remarks, "If people truly understood the crimes committed to the people and our world just by perverting our communication systems and hampering most software innovation so we could buy iPhone they would be appalled." P18 feels frustrated with the control afforded by capital, "It is very difficult to look at our current world and have hope that things will get better unless it's entirely by accident or more largely benefits the wealthy."

Others appeal to the collective, advocating for empathy. Three (3) respondents express that the mitigation of misinformation is everyone's responsibility. P13 explains, "Social media platforms already have a difficult time keeping up with bots that spread misinformation, so let's help the platforms ourselves by spreading the truth." P7 shares a similar sentiment, asserting that objectivity when mitigating misinformation is crucial: "Misinformation and misinformed people can be of any political ideology, political warfare should never happen when educating people." P5 believes that "information should be free, as long as it is correct and spread without an agenda." Finally, P9 re-asserts their ideology stating that "free men don't ask."

[This space left intentionally blank for formatting purposes]

4.1.11 Question 11

If you would like to share some demographic information, please do - we do not require it, but it will help us better contextualize your effort and story.

Gender Identity was the only fully disclosed demographic data, either via social media profiles or directly volunteered; thirteen (13) respondents did not disclose a comprehensive demographic makeup and nine (9) did not respond to the query at all. Gender distribution skews male-identifying at fifteen (15), with nine (9) female-identifying and one (1) respondent identifying as non-binary. Eleven (11) participants disclosed their age; eight (8) of whom identify as Millennial (23-38 years old) and three (3) identify as Generation X. Eight (9) respondents identified as from the United States, two (2) identified as from the United Kingdom, one (1) from Canada and one (1), P12, simply stated "Europe".

FIGURE 9. Q11 TOTALS

MALE-IDENTIFYING	16
FEMALE-IDENTIFYING	8
NON-BINARY	1
MILLENIAL	8
GEN-X	3
NORTH AMERICA	8
EUROPE	1
UNITED KINGDOM	2
CANADA	1

4.2 Observed Operations: JonathanData1 Vs. Citizen Labs

In this section I evaluate one of the operations that inspired this research, the tale of JonathanData1 versus Citizen Labs, an Amnesty International affiliate. Jonathan releases a report to a researcher social media site named ResearchGate, debunking Citizen Lab's report detailing the domestic spying perpetrated by the Spanish government on the Catalan Separatist movement using mobile spyware, purchased from NSO group. In the report he unleashes a scathing rebuke of Citizen Lab's perceived biases and lack of qualifications to investigate the matter, declaring himself the world's premier mobile spyware expert. Jonathan's report is debunked, many times, by a number of journalists and cybersecurity experts.

1. **Primary Diffusion Vector**: Johnathan Lee Villareal aka Johnathan Scott via Twitter (Primary), Substack (Secondary) and ResearchGate (payload repository).
2. **Original Payload:** Review of Catalangate by Johnathan Scott

3. **Claim**: The Spanish government did NOT use Pegasus and Candiru spyware on Catalan separatists, contrary to an April 2021 report published by CitizenLabs and Amnesty International. These organizations are fraudulent for publishing their report.
4. **Classification:** Where [brackets] denote classification, using the Sharevski Folk Model[46] as a Primary Classifier and the Zannetou Et Al[58] models as a sub-category.

4.2.1 Classification

This is claim is [*Inherently Fallacious*] and it creates a [*Conspiracy Theory*] implicating nefarious intent behind the collusion of CitizenLabs and Amnesty International in their effort to expose the Spanish government's spying on Catalan separatists. More importantly, the accusation of conspiracy is used as a rhetorical cudgel of sorts. Any detractor is met with relentless ad-hominem attacks and demands for either clarification or rendering of credentials. Johnathan uses force escalation and frequency to subdue his detractors, where possible.

4.2.2 Actor Profile

Johnathan's academic credibility in the public eye was manufactured by several [*out-of-context narratives*] weaved in his honor by CNN, the New York Post and Wall Street Journal. Jonathan fallaciously claimed that the Beijing Olympics App that all participants were required to use was spying on everyone. Jonathan exploits the lack of cybersecurity expertise held by most mainstream journalists by boasting his (revoked) status as #1 contributor on HackerOne and premier mobile malware forensic researcher.

4.2.3 Opposition Profile

That was until he ran afoul of the larger Information Security community, notably Lukasz "Maldr0id", a Google Android developer and Runa Sandvik, former New York Times Senior Director of Information Security. Both Runa and Lukasz successfully deployed [*debunking*] operations using [*direct refutation*] and, eventually, [*emotionally charged*] shared experiences, outing Jonathan as actively harassing them.

4.2.4 Counter-operations

While direct refutation has largely killed Jonathan's reputation with the press and the larger infosec community, he has repeated this operation twice more. The pattern is exactly the same, in all three operations. Jonathan targets a humanitarian cause, usually one where the Pegasus Mobile Spyware is implicated as a weapon of the oppressor, publishes a fraudulent research paper to ResearchGate, followed by a post to Substack with a summary and sustained by rapid, frequent Tweets on the topic. When met with refutation or rebuttal, he asserts his credibility as dominant and the opposition as unqualified to interact.

4.2.5 Actor Retaliation

During the Anti-Catalangate campaign mentioned here, this tactic wore down his detractors one after another until a few banded together to reach out to his academic program to inform them of his behavior. Once he was ejected from his Doctoral program, he adopted the language of the western culture wars. Anyone who disagreed with him was "woke" and all assaults were the result of "cancel culture." This has attracted a host of Twitter followers who are recently un-banned or have a history of interacting with Alt-right or MAGA-associated accounts. Most of the supporters he gained by adopting this language are not security experts and do not question his credentials. His attacks on journalists and civil rights watchdogs are [bias affirming] for those who consider social justice or feminism a societal ailment. In this model, Jonathan doesn't need to deal in facts, he just needs to be [part of the in-group].

[This space left intentionally blank for formatting purposes]

Chapter 5: Conclusions

5.1 Research Question 1

How do people who consider themselves as active operational participants in cyberspace conceptualize misinformation?

Eleven (11) respondents view misinformation as external to themselves, only three (3) acknowledge having consumed misinformation. Eight (8) respondents conceptually oppose the diffusion of misinformation into online social networks. However, ten (10) respondents claim to be actively operating to oppose misinformation diffusion, with four (4) claiming to having been inactive. The surveyed respondents largely acknowledge misinformation's existence as problematic for social spaces online but, only 56% have engaged in counter-operations.

Cyber-actors' detailed mental models of misinformation are highly varied but, most view it an external or direct threat to their activities online. When defining the concept, some categorize misinformation as either External Propaganda or Biased information. The prevailing sentiment is that misinformation's ever-presence is the result of human bias and a failure of education. Where that education lies, however, is variable; be it a nebulous system or peer-led learning, it is an external resource that is found to be lacking in the eyes of cyber-actors.

5.2 Research Question 2

What are the tools, tactics, and procedures these "cyber-actors" employ in dealing with any form of falsehoods or questionable information in cyberspace?

Most models are based on intrapersonal or small-scale operations, respondents notably did not discuss persistence or how to ascertain efficacy. While respondents seem wholly capable of planning, pump-priming and, occasionally, offensive tactics, their individual operational models are incomplete. Furthermore, the total lack of responses in the "persistence" and "assess effectiveness" categories of the DISARM BLUE framework is telling. Where the operational mindset of a cyber actor, especially hackers & cyber-professionals is valuable, it lacks an understanding of macro-scale psycho-social systems present in online social networks. This is

not lost on some cyber-actors, there is mention of their lack of understanding on how to fully execute or assess the success of a counter-misinformation operation. If the individual actors in this group were to collaborate, they are likely to concoct a mostly complete operation without a playbook. Figure 10 on the next page (48) illustrates this.

5.3 Research Question 3

What is the best response against misinformation in the view of these "cyber-actors"?

The theme of education re-occurs throughout the survey: educate your peers, educate your social circles and teach media literacy at all levels. Cyber-actors also underscore the importance expose the leading actors or "grifters", as some put it. Doing so within the systems bad actors are using to diffuse misinformation is a key component of an aggressive direct-action campaign. Conversely, many ask us to push back on misinformation with empathy, in one-on-one interactions, urging the reader to understand the consumer's context. Further, cyber actors feel we must put pressure on legislators, advertisers, and social media platform owners to participate in counter-misinformation activity. Moreover, the public should condemn the continued diffusion of misinformation in their personal spaces, armed with the truth. Finally, P24 puts it best, urging the reader to internalize that "The act of resistance and fighting is more important than winning."

[This space left intentionally blank for formatting purposes]

Figure 10. DISARM Blue Framework - Cyber Actor Preferred TTPs									
TA01:	**TA02:**	**TA05:**	**TA06:**	**TA07:**	**TA08:**	**TA09:**	**TA11:**	**TA12:**	**TA15:**
Plan Strategy	Plan Objectives	Microtarget	Develop Content	Select Channels and Affordances	Conduct Pump Priming	Deliver Content	Persist in the Information Environment	Assess Effectiveness	Establish Social Assets
C00017:	C00029:	C00216:	C00032:	C00105:	C00119:	C00124:			C00052:
Repair broken social connections	Create fake website to issue counter narrative and counter narrative through physical merchandise	Use advertiser controls to stem flow of funds to bad actors	Hijack content and link to truth-based info	Buy more advertising than misinformation creators	Engage payload and debunk.	Don't feed the trolls			Infiltrate platforms
C00019:	C00030:	C00130:	C00080:		C00154:				C00133:
Reduce effect of division-enablers	Develop a compelling counter narrative (truth based)	Mentorship: elders, youth, credit. Learn vicariously.	Create competing narrative		Ask media not to report false information				Deplatform Account*
C00153:	C00011:		C00087:		C00136:				
Take pre-emptive action against actors' infrastructure	Media literacy. Games to identify fake news		Make more noise than the disinformation		Microtarget most likely targets then send them countermessages				
C00111:					C00113:				
Reduce polarization by connecting and presenting sympathetic renditions of opposite views					Debunk and defuse a fake expert / credentials.				
C00170:					C00114:				
elevate information as a critical domain of statecraft					Don't engage with payloads				
					C00115:				
					Expose actor and intentions				

Chapter 6: Discussions

6.1 Ethical Consideration - On cyber-vigilantism

This work is not presented as a justification or endorsement of hacktivism, hacking, vigilantism, doxing, or any offensive security tactic. The actors who disclosed their activities were not asked to justify their activity, nor were they expected to disclose specific operations. This work is not intended to create a body of evidence that provides ethical or analytical justification for any actions disclosed in the cyber-actors' own words. I caution the reader to consider the material and cognitive harm some counter-operations may have on unwitting misinformation consumers. It is not my intention to provide a primer for the moralizing of vigilantism if it were to apply to misinformation operations. As such, I present my own experience, growing close to a hacktivist and failing to consider the externalities that influenced this cyber-actor's journey. Let this be a word of caution to the reader, where an actor's agenda may be bias affirming, their material conditions and practical externalities may be in direct conflict with your own moralizing.

During the study, I had the good fortune to become close to a controversial threat actor presenting themselves as a hacktivist known as AgainstTheWest. The group readily identified as hacktivist, announcing their adversaries as Russia, Iran, and China to anyone who would listen. The group has now disbanded and is known to be defunct but, the researcher was able to maintain contact between December 2021 and July 2022. Researchers took an early opportunity to gain access to hacktivists and potentially embed with a group.

They presented a convincing facade, proclaiming they were prosecuting the injustices of unjust regimes by exploiting and leaking data from NATO-adversary government contractors. Proudly declaring in several hacker communities that the west should not fear China and Russia anymore, they drew derision from many and suspicion from most. A very appealing narrative, the researcher was near convinced they were a rare breed of hacktivist, worthy of additional study. After contacting their most vocal member, they spoke at length to the researcher about Muslim oppression in mainland China,

claiming their groups actions against state affiliated entities was retribution for the injustices suffered under Chinese government.

A self-sufficient vengeful group standing against state-level actors was a compelling narrative and created significant discourse in the communities they touched as well as within development of the presented body of work. What's more, the narrative suited the researcher's model of the hacktivist, a group of outsiders against the world. The level of risk was palpable for this group, and they freely admitted to it; during their interview for this study, the actor admitted that they were afraid for their life. Insisting that their work against China and Russia would get them assassinated or worse before their mission was complete. Throwing themselves at activism seemed to be their attempt at leaving a legacy before dying potentially very young.

As we now know, one of the communities where this actor made the most noise was compromised by US and international law enforcement for some time and [RaidForums citation needed]. What's more, the actor's now defunct telegram channel confesses to being a state sponsored actor [ATW telegram citation needed]. Hacktivism appeared to be a cover for state-level cyberwarfare, a common accusation levied at this actor amongst their peers. Some weeks on, the researcher was contacted by a purported associate of the defunct threat actor; the previous contact's prognostication that the work could kill them had come to pass. Our first survey participant was allegedly deceased thanks to terminal brain cancer. The new contact implied that group's work was a subsidiary to a US Intelligence Community operational unit and our contact's participation in cyberwarfare was allowed and supported specifically because of their short time left on the planet.

Disillusioning, it begged the question: how much cyber activism is genuinely grassroots? How many seemingly independent actors were exploited and treated as disposable by larger organizations to sow disorder in the systems of an adversary? Is it worth considering in the scope of this study? Could it safely have vetted for truth, considering the potential stakes? Is hacktivism limited to ideologies or upheaval or can it be broader and more dangerous? What differentiates this kind of righteousness from

cyber-terrorism? In an age of bug bounties, why not also cyber vigilantes; working as guerillas for more global initiatives?

6.2 Challenges – Trust Relationships

Talking to the cyber actors in the field presented a few unexpected issues, worth noting here. Primarily, the development of trust relationships and the implications of maintaining them inside a complex social ecosystem. As a relatively new actor in the social spaces, I targeted during this study, my presence was justifiably suspect. In the immediate, I was fortunate enough to understand local vernacular, either by listening in or from previous exposure. In the medium term, however, I had no alibi, my credibility as an academic and not a "Fed" was largely reliant on my .edu email address.

Those suspect of my intentions used familiar methods to dislodge me from their communities. One actor chose to spread an unsavory rumor to their network about my past, another spread my personal information to the very community I was attempting to ingratiate myself with. On that note, I was rewarded for my good faith, each time affirmed by would-be respondents or new connections that my efforts were worthy. Initially, I suspected I should have performed this research under a false persona, to protect myself. In retrospect and, having conversed with some respondents after-the-fact, it was my earnest effort and vulnerability that gained the trust and respect of the participants of this study. So, I would offer this only as a word of caution to future researchers: there will always be bad actors willing to tear you down for your troubles but, research worth doing is almost never without risk or difficulty.

Bibliography

[1] Yochai Benkler, Robert Faris, and Hal Roberts. *Network propaganda: Manipulation, disinformation, and radicalization in American politics*. Oxford University Press, 2018.

[2] Davide Beraldo. “Unfolding #Anonymous on Twitter: The networks behind the mask.” In: *First Monday* 27.1 (2023/01/20 2022).

[3] Jack Brewster et al. *Beware the ‘New Google:’ TikTok’s Search Engine Pumps Toxic Misinformation To Its Young Users*. 2022. url: https://www.newsguardtech. com/misinformation-monitor/september-2022/.

[4] Victoria Carty. *Social Movements and New Technology*. Taylor and Francis, 2018.

[5] Gabriella Coleman. *Hacker, hoaxer, whistleblower, spy: The many faces of Anonymous*. Verso books, 2014.

[6] Ellen Cornelius. “Anonymous Hacktivism: Flying the Flag of Feminist Ethics for the Ukraine IT Army.” In: (2022).

[7] John Corner. *Fake news, post-truth, and media–political change*. 2017.

[8] Glenn Diesen. “Conclusion: Anti-Russian Propaganda of a West in Relative Decline.” In: Singapore: Springer Nature Singapore, 2022, pp. 255–258.

[9] Abbas Ehsanfar and Mo Mansouri. “Incentivizing the dissemination of truth versus fake news in social networks”. In: *2017 12th System of Systems Engineering Conference (SoSE)*. 2017, pp. 1–6. doi: 10.1109/SYSOSE.2017.7994981.

[10] Luca Follis and Adam Fish. “3 When to Hack”. In: *Hacker States*. 2020, pp. 73–111.

[11] DISARM Foundation. *What is the DISARM framework?* https://www.disarm.foundation/framework. Accessed: March 5, 2023. 2023.

[12] Deen Freelon, Charlton D McIlwain, and Meredith Clark. “Beyond the hashtags:# Ferguson,# BlackLivesMatter, and the online struggle for offline justice”. In: *Center for Media & Social Impact, American University, Forthcoming* (2016).

[13] Jordana J. George and Dorothy E. Leidner. “From clicktivism to hacktivism: Understanding digital activism”. In: *Information and Organization* 29.3 (2019), p. 100249.

[14] Paolo Gerbaudo. “From cyber-autonomism to cyber-populism: An ideological history of digital activism”. English. In: *tripleC: Communication, Capitalism & Critique* 15.2 (May 2017), pp. 477–489. issn: 1726-670X.

[15] Ritu Gill. *What is Open-Source Intelligence?* https://www.sans.org/blog/what-isopen-source-intelligence/. Accessed: March 5, 2023. 2023.

[16] Claire Goforth. *‘Anonymous’ hackers have a message for Texas abortion ‘snitch’ sites: We’re coming for you*. 2021. url: https://www.dailydot.com/debug/anonymous-hactivists-texas-abortion-ban-operation-jane/.

[17] Luke Goode. “Anonymous and the Political Ethos of Hacktivism”. In: *Popular Communication* 13.1 (2015), pp. 74–86.

[18] Jason Hannan. “Trolling ourselves to death? Social media and post-truth politics”. In: *European Journal of Communication* 33.2 (2018), pp. 214–226.

[19] Masayuki Hatta. “Cowboys and the Eternal September Transfiguration of hacker aesthetics”. In: *Annals of Business Administrative Science* (2021), 0210923a.

[20] Laura Illia. “Passage to cyberactivism: how dynamics of activism change”. In: *Journal of public affairs.* 3.4 (2003-11).

[21] Jane Im et al. “Still out There: Modeling and Identifying Russian Troll Accounts on Twitter”. In: *12th ACM Conference on Web Science*. WebSci ’20. Southampton, United Kingdom: Association for Computing Machinery, 2020, pp. 1–10. isbn: 9781450379892. doi: 10.1145/3394231.3397889. url: https://doi.org/10. 1145/3394231.3397889.

[22] Leanna Ireland. “We are all (not) Anonymous: Individual- and country-level correlates of support for and opposition to hacktivism”. In: *New Media & Society* 0.0 (0), p. 14614448221122252.

[23] Keenan Jones, Jason R. C. Nurse, and Shujun Li. “Behind the Mask: A Computational Study of Anonymous’ Presence on Twitter”. In: *Proceedings of the International AAAI Conference on Web and Social Media* 14.1 (May 2020), pp. 327–
338.

[24] Andreas Jungherr, Gonzalo Rivero, and Daniel Gayo-Avello. *Retooling Politics: How Digital Media Are Shaping Democracy*. Cambridge University Press, 2020. doi: 10. 1017/9781108297820.

[25] Vasileios Karagiannopoulos. “A Short History of Hacktivism: Its Past and Present and What Can We Learn from It”. In: ed. by Tim Owen and Jessica Marshall. Cham: Springer International Publishing, 2021, pp. 63–86.

[26] Athina Karatzogianni. *Firebrand waves of digital activism 1994-2014: The rise and spread of hacktivism and cyberconflict*. Springer, 2015.

[27] Steven Levy. *Hackers: Heroes of the Computer Revolution - 25th Anniversary Edition*. 1st. O'Reilly Media, Inc., 2010. isbn: 1449388396.

[28] Alexander J Lindvall. "Political hacktivism: doxing & the first amendment". In: *Creighton L. Rev.* 53 (2019), p. 1.

[29] Ioana Literat, Lillian Boxman-Shabtai, and Neta Kligler-Vilenchik. "Protesting the Protest Paradigm: TikTok as a Space for Media Criticism". In: *The International Journal of Press/Politics* 0.0 (0), p. 19401612221117481.

[30] Clare Llewellyn et al. "Russian Troll Hunting in a Brexit Twitter Archive". In: *Proceedings of the 18th ACM/IEEE on Joint Conference on Digital Libraries*. JCDL '18. Fort Worth, Texas, USA: Association for Computing Machinery, 2018, pp. 361– 362. isbn: 9781450351782. doi: 10.1145/3197026.3203876. url: https://doi. org/10.1145/3197026.3203876.

[31] Martha McCaughey and Michael D Ayers. *Cyberactivism: Online activism in theory and practice*. Psychology Press, 2003.

[32] Ty McCormick. "Anthropology of an Idea Hacktivism". In: *Foreign Policy* 200 (May 2013), pp. 24–25.

[33] Stefania Milan. "Hacktivism as a radical media practice". In: *The Routledge companion to alternative and community media*. Routledge, 2015, pp. 550–560.

[34] Stefania Milan. *Social movements and their technologies: Wiring social change*. Springer, 2013.

[35] Mahdi M. Najafabadi and Robert J. Domanski. "Hacktivism and distributed hashtag spoiling on Twitter: Tales of the #IranTalks". In: *First Monday* 23.4

(Apr. 2018). doi: 10.5210/fm.v23i4.8378. url: https://firstmonday.org/ojs/index.php/ fm/article/view/8378.

[36] Mark Rolfe. “Hacker: Creating the Narrative of the Digital Robin Hood”. In: *The Reinvention of Populist Rhetoric in The Digital Age: Insiders & Outsiders in Democratic Politics*. Springer Singapore, 2016, pp. 135–164. doi: 10.1007/978-981-102161-9_6. url: https://doi.org/10.1007/978-981-10-2161-9_6.

[37] Marco Romagna. “Hacktivism: Conceptualization, Techniques, and Historical View”. In: ed. by Thomas J. Holt and Adam M. Bossler. Cham: Springer International Publishing, 2020, pp. 743–769.

[38] Rodrigo Sandoval-Almazan and J. Ramon Gil-Garcia. “Towards cyberactivism 2.0? Understanding the use of social media and other information technologies for political activism and social movements”. In: *Government Information Quarterly* 31.3 (2014), pp. 365–378.

[39] Madelyn R Sanfilippo, Shengnan Yang, and Pnina Fichman. “Managing online trolling: From deviant to social and political trolls”. In: *50th Annual Hawaii International Conference on System Sciences, HICSS 2017*. IEEE Computer Society. 2017, pp. 1802–1811.

[40] Saiph Savage, Andres Monroy-Hernandez, and Tobias H¨ollerer. “Botivist: Calling Volunteers to Action Using Online Bots”. In: *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. CSCW ’16. San Francisco, California, USA: Association for Computing Machinery, 2016, pp. 813– 822. isbn: 9781450335928. doi: 10.1145/2818048.2819985. url: https://doi. org/10.1145/2818048.2819985.

[41] Mark Scott. *The shit-posting, Twitter-trolling, dog-deploying social media army taking on Putin one meme at a time*. 2022. url: https://www.politico.eu/article/nafo-doge-shiba-russia-putin-ukraine-twitter-trolling-social-mediameme/.

[42] Philip Serracino-Inglott. “Is it OK to be an Anonymous?” In: *Ethics & Global Politics* 6.4 (2013), p. 22527.

[43] Lanyu Shang et al. “A Multimodal Misinformation Detector for COVID-19 Short Videos on TikTok”. In: *2021 IEEE International Conference on Big Data (Big Data)*. 2021, pp. 899–908. doi: 10.1109/BigData52589.2021.9671928.

[44] Filipo Sharevski et al. *“Gettr-ing” User Insights from the Social Network Gettr*. https://truthandtrustonline.com/wp-content/uploads/2022/10/TTO_2022_proceedings.pdf. Boston, MA, 2022.

[45] Filipo Sharevski et al. “Folk Models of Misinformation on Social Media”. In: *NDSS Symposium* (2023).

[46] Filipo Sharevski et al. “Meaningful Context, a Red Flag, or Both? Preferences for Enhanced Misinformation Warnings Among US Twitter Users”. In: *Proceedings of the 2022 European Symposium on Usable Security*. EuroUSEC ’22. https://doi.org/ 10.1145/3549015.3555671. Karlsruhe, Germany: Association for Computing Machinery, 2022, pp. 189–201. isbn: 9781450397001. doi: 10.1145/3549015.3555671.

[47] Filipo Sharevski et al. “VoxPop: An Experimental Social Media Platform for Calibrated (Mis)Information Discourse”. In: *New Security Paradigms Workshop*. NSPW ’21. Virtual Event, USA: Association for Computing Machinery, 2021,

pp. 88–107. isbn: 9781450385732. doi: 10.1145/3498891.3498893. url: https://doi.org/
10.1145/3498891.3498893.

[48] Micah L Sifry. *WikiLeaks and the Age of Transparency*. OR Books, 2011.

[49] Edward Snowden. *Permanent record*. Pan Macmillan, 2019.

[50] Tom Sorell. “Human Rights and Hacktivism: The Cases of Wikileaks and Anonymous”. In: *Journal of Human Rights Practice* 7.3 (Sept. 2015), pp. 391–410.

[51] Kevin F Steinmetz. “Hacking And Hacktivism”. In: *Shades of Deviance: A Primer on Crime, Deviance and Social Harm* 19 (2022).

[52] Justus Uitermark. “Complex contention: analyzing power dynamics within Anonymous”. In: *Social Movement Studies* 16.4 (2017), pp. 403–417.

[53] Jared M Wright et al. “Drive-By Solidarity: Conceptualizing the Temporal Relationship between# BlackLivesMatter and Anonymous’s# OpKKK”. In: *Contention* 10.2 (2022), pp. 25–55.

[54] Liang Wu et al. “Misinformation in Social Media: Definition, Manipulation, and Detection”. In: 21.2 (2019). url: https://kdd.org/exploration_files/8._CR.10.Misinformation_in_social_media_-_Final.pdf.

[55] Savvas Zannettou et al. “Characterizing the Use of Images in State-Sponsored Information Warfare Operations by Russian Trolls on Twitter”. In: *Proceedings of the International AAAI Conference on Web and Social Media* 14.1 (May 2020), pp. 774–
785. doi: 10.1609/icwsm.v14i1.7342. url: https://ojs.aaai.org/index.php/ICWSM/article/view/7342.

[56] Savvas Zannettou et al. “On the Origins of Memes by Means of Fringe Web Communities”. In: *Proceedings of the Internet Measurement Conference 2018*. IMC ’18. Boston, MA, USA: Association for Computing Machinery, 2018, pp. 188–202. isbn: 9781450356190. doi: 10.1145/3278532.3278550. url: https://doi.org/10. 1145/3278532.3278550.

[57] Savvas Zannettou et al. “The Web of False Information: Rumors, Fake News, Hoaxes, Clickbait, and Various Other Shenanigans.” In: *J. Data and Information Quality* 11.3 (May 2019). issn: 1936-1955. doi: 10.1145/3309699. url: https://doiorg.ezproxy.depaul.edu/10.1145/3309699.

[58] Savvas Zannettou et al. “Who Let The Trolls Out? Towards Understanding StateSponsored Trolls.” In: *Proceedings of the 10th ACM Conference on Web Science*. WebSci ’19. New York, NY, USA: Association for Computing Machinery, 2019, pp. 353–362.