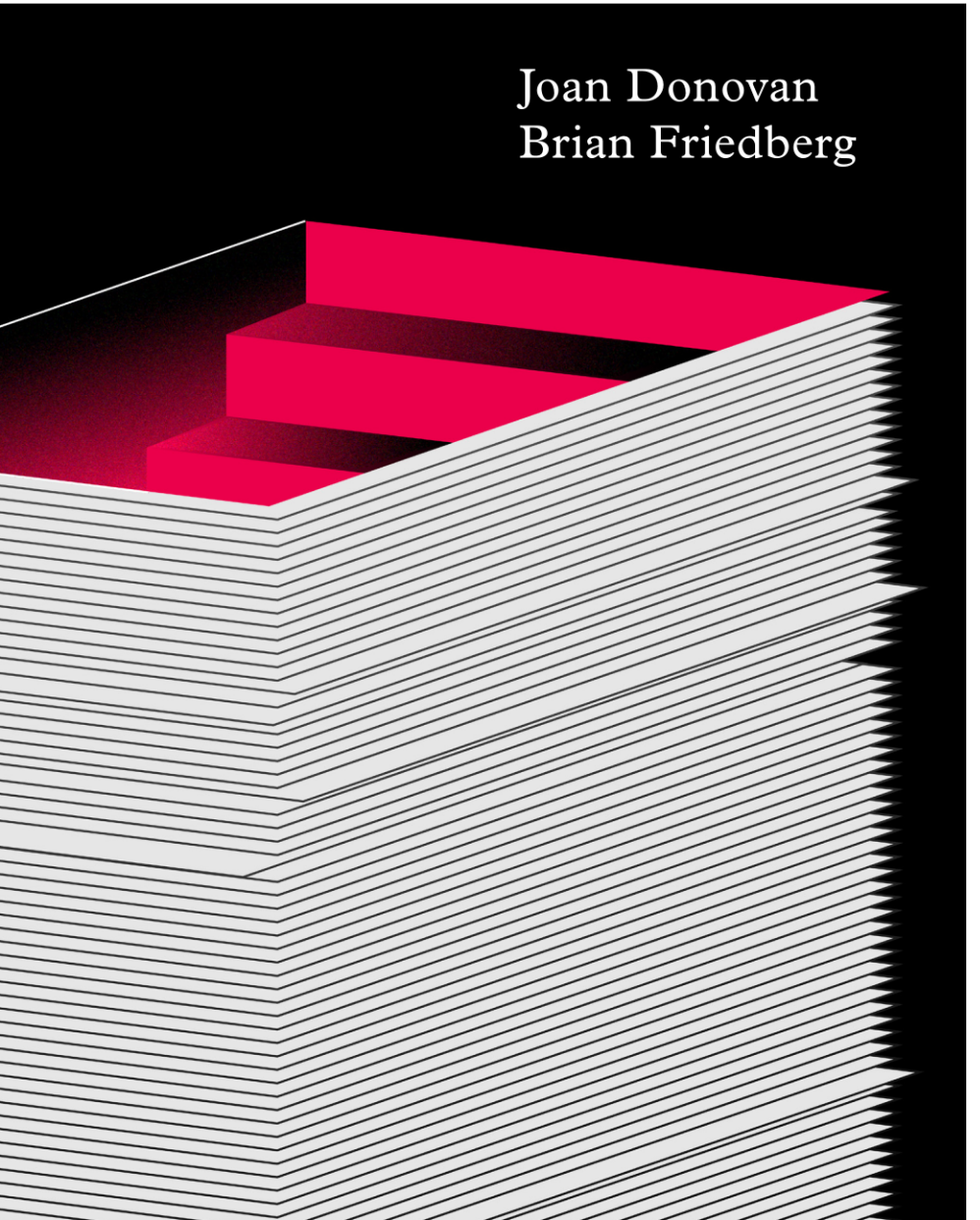
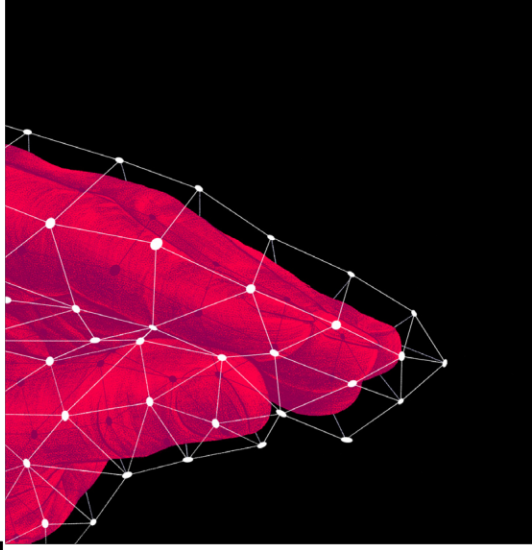
DATA&
SOCIETY

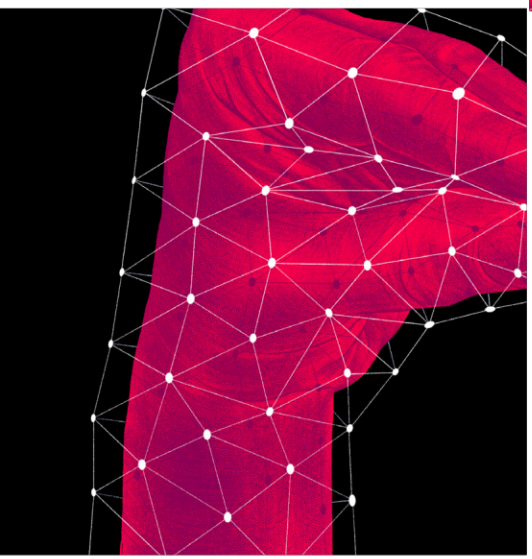
SOURCE HACKING

MEDIA MANIPULATION IN PRACTICE

Joan Donovan
Brian Friedberg





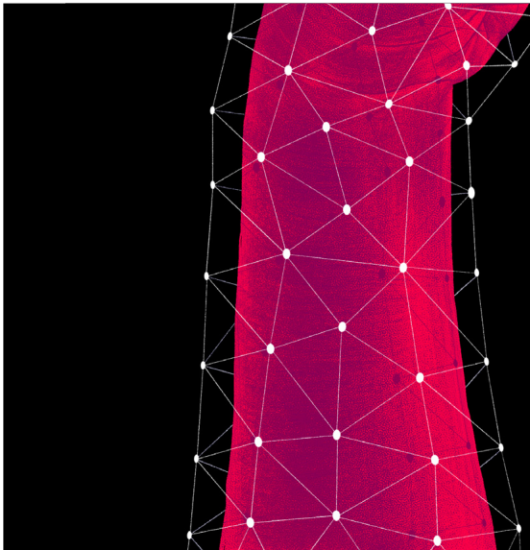


Author: Joan Donovan; Director of the Technology and Social Change Research Project, Harvard Kennedy School, PhD, 2015, Sociology and Science Studies, University of California San Diego.

Author: Brian Friedberg; Senior Researcher, Technology and Social Change Research Project, Harvard Kennedy School; MA, 2010, Cultural Production, Brandeis University.

This report is published under Data & Society's Media Manipulation research initiative; for more information on the initiative, including focus areas, researchers, and funders, please visit https://datasociety.net/research/media-manipulation

CONTENTS



EXECUTIVE SUMMARY

In recent years there has been an increasing number of **online manipulation campaigns** targeted at news media. This report focuses on a subset of manipulation campaigns that rely on a strategy we call **source hacking**: a set of techniques for hiding the sources of problematic information in order to permit its circulation in mainstream media. Source hacking is therefore an indirect method for targeting journalists—planting false information in places that journalists are likely to encounter it or where it will be taken up by other intermediaries.

Across eight case studies, **we identify the underlying techniques of source hacking to provide journalists, news organizations, platform companies, and others with a new vocabulary for describing these tactics**, so that terms such as "trolling" and "trending" do not stand in for concerted efforts to pollute the information environment. In this report, we identify four specific techniques of source hacking:

1. **Viral Sloganeering:** repackaging reactionary talking points for social media and press amplification
2. **Leak Forgery:** prompting a media spectacle by sharing forged documents
3. **Evidence Collages:** compiling information from multiple sources into a single, shareable document, usually as an image
4. **Keyword Squatting:** the strategic domination of keywords and sockpuppet accounts to misrepresent groups or individuals

These four tactics of source hacking work because networked communication is vulnerable to many different styles of attack and finding proof of coordination is not easy to detect. Source hacking techniques complement

each other and are often used simultaneously during active manipulation campaigns. These techniques may be carefully coordinated but often rely on partisan support and buy-in from audiences, influencers, and journalists alike.

We illustrate these techniques with case studies taken from 2016–2018, and with a specific focus on the manipulation of American politics. We end by offering a set of suggestions and new concepts for those attempting to identify the operations of manipulation campaigns or to respond to breaking news events:

- We advise journalists to seek out an **abundance of corroborating evidence** when reporting on the actions of social media accounts, and whenever possible, verify the identity of account holders.

- We suggest that **newsrooms invest more resources in information security**, including creating a position or desk to vet chains of evidence through analysis and verification of metadata for evidence of data craft.

- We argue that **platform companies must label manipulation campaigns when they are identified** and provide easier access to metadata associated with accounts.

INTRODUCTION

In recent years, there has been an increasing number of online **manipulation campaigns** targeted at news media. The goals of manipulation campaigns can vary widely, but they all rely on communication platforms to respond in real time to breaking "media spectacles" or, sometimes, to anticipate or even generate such spectacles.[1] This report focuses on a subset of manipulation campaigns that rely on a strategy we call **source hacking**. Source hacking is a set of techniques for hiding the sources of problematic information in order to permit its circulation in mainstream media.[2] Source hacking is therefore an indirect method for targeting journalists—planting false information in places that journalists are likely to encounter it, or where it will be taken up by other intermediaries.

1 Douglas Kellner, *Media Spectacle* (London; New York: Routledge, 2003).

2 The concept of cloaked websites, which operate like phishing campaigns, originates with research related to white supremacists' use of the internet. See: Jessie Daniels, *Cyber Racism: White Supremacy Online and the New Attack on Civil Rights* (Lanham, Md: Rowman & Littlefield Publishers, 2009). Md: Rowman & Littlefield Publishers, 2009.

In this report, we identify four specific techniques of source hacking: **viral sloganeering, leak forgery, evidence collaging,** and **keyword squatting.** Using these categories to describe the work of manipulation campaigns can increase the specificity with which we track media manipulation. We illustrate these techniques with case studies taken from 2016–2018, and with a specific focus on the manipulation of American politics.[3] We end by offering a set of suggestions and new concepts for those attempting to identify the operations of manipulation campaigns or to respond to breaking news events.[4]

3 Media manipulation campaigns are drawn from the local political opportunities aided by the organization of the state, technology companies, and news media. We do not know the extent to which the tactics we observe translate into other contexts. For grounded empirical investigations, we recommend the following research: India (Chaturvedi, Swati. 2016. I Am a Troll: Inside the Secret World of the BJP's Digital Army. New Delhi, India: Juggernaut Publication). Philippines (Ong, J. C., and J. V. A. Cabanes. 2018. "Architects of Networked Disinformation: Behind the Scenes of Troll Accounts and Fake News Production in the Philippines." Monograph. February 9, 2018. http://newtontechfordev.com/wp-content/uploads/2018/02/ARCHITECTS-OF-NETWORKED-DISINFORMATION-FULL-REPORT.pdf.), Russia, (Pomerantsev, Peter. 2015. Nothing Is True and Everything Is Possible: The Surreal Heart of the New Russia. Reprint edition. PublicAffairs.), Mexico (Rosa, Raúl Magallón. 2019. UnfakingNews: Cómo combatir la desinformación. Edición: edición. Grupo Anaya Publicaciones Generales).

4 We acknowledge the potential impact of amplifying manipulation efforts through research. Scholars Ryan Millner and Whitney Phillips have explored the ways in which granting additional "oxygen" to these actors helps spread their message. While identifying proven techniques of media manipulators, we avoid providing insight into how manipulation could be done more effectively. Additionally, we are not attempting to highlight individual manipulators or grant them platforms. Instead, this document provides a historical context for future manipulation campaigns and recommendations for reporting on extremism. See: Ryan Milner, "Hacking the Social: Internet Memes, Identity Antagonism, and the Logic of Lulz," *The Fibreculture Journal*, no. 22 (2013), http://twentytwo.fibreculturejournal.org/fcj-156-hacking-the-social-internet-memes-identity-antagonism-and-the-logic-of-lulz/; Whitney Phillips, "The Oxygen of Amplifcation: Better Practices for Reporting on Extremists, Antagonists, and Manipulators Online" (New York: Data & Society Research Institute, May 2018).

WHAT IS SOURCE HACKING?

Source hacking is a versatile set of techniques for feeding false information to journalists, investigators, and the general public during breaking news events or across highly polarized wedge issues.[5] Specifically, source hacking exploits situations and technologies to obscure the authorship of false claims. Across eight case studies, we identify the underlying techniques of source hacking to provide journalists, news organizations, platform companies, and others with a new vocabulary for describing these tactics, so that terms such as "trolling" and "trending" do not stand in for concerted efforts to pollute the information environment.

Manipulators must be skillful content creators, adept at creating persuasive audiovisual materials in time-sensitive scenarios. Persuasive memes, convincing fake articles, and heavily edited video are common media spread during a manipulation campaign. Content is often workshopped in private communications or on forums. Merging the techniques of marketing and political propaganda, successful source hacking materials are persuasive text, images, or video that do not require a legitimate author to establish authenticity.

Manipulators demonstrate skill both in the crafting of a message and the metadata of the objects used as

5 Wedge issues are contested politicized positions around identity, authority, and justice, often centered on the distribution of rights and representation. The support for, and media representation of, civil rights organizing has historically been a wedge issue in the United States. See, for instance: Gyung-Ho Jeong et al., "Cracks in the Opposition: Immigration as a Wedge Issue for the Reagan Coalition," *American Journal of Political Science* 55, no. 3 (2011): 511–25.

evidence. *Data craft*, as defined by Amelia Acker (2018), is a set of practices that "create, rely, or even play with the proliferation of data on social media by engaging with new computational and algorithmic mechanisms of organization and classification."[6] Media manipulators create campaign materials with knowledge of technological and cultural vulnerabilities, taking advantage of platform design to amplify persuasive content. Careful attention to data craft is the foundation of a successful manipulation campaign, as manipulators create images, videos, and documents, carefully working around systems like spam detection, circumventing platform companies' terms of service and trust and safety teams. Individual actors use sockpuppet accounts, botnets, and social media influencers to create trending opportunities and saturate hashtags with original content. Careful data crafting allows manipulators to create forgeries that appear legitimate and original, obscuring their origins with appeals to authenticity and using metrics as a form of legitimation.[7]

The following section presents eight case studies of manipulation campaigns from 2016–2018. These campaigns vary in many ways. Some were quick responses to breaking news coverage, while others were deliberate attempts to *create* new coverage. Some events, like "It's OK To Be White," were particularly campaign-like and involved the careful design of a slogan by a small number of manipulators. Others, like Pizzagate, may have had periods of campaign-like organization, but were also the result of long, messy collaboration across platforms and groups. Despite these differences, all of these campaigns relied on some form of source hacking. That is, the campaigns'

6 Amelia Acker, "Data Craft" (Data & Society Research Institute, November 5, 2018), https://datasociety.net/output/data-craft/.

7 Acker developed an eight-step chart for spotting manipulation on social media that all journalists, researchers, and technologists can use to identify red flags. See: https://datasociety.net/wp-content/uploads/2018/11/DS_Data_Craft_Manipulation_of_Social_Media_Metadata_Infographic2.pdf.

success relied specifically on hiding the real sources of information from an eventual mainstream audience.

The four techniques of source hacking we identify are:

1. Viral Sloganeering: repackaging reactionary talking points for social media and press amplification

2. Leak Forgery: prompting a media spectacle by sharing forged documents

3. Evidence Collages: compiling information from multiple sources into a single, shareable document, usually as an image

4. Keyword Squatting: the strategic domination of keywords and sockpuppet accounts to misrepresent groups or individuals

1. VIRAL SLOGANEERING CASE STUDIES:

- Jobs Not Mobs
- It's OK To Be White

Viral sloganeering *is a process of crafting divisive cultural or political messages in the form of short slogans and propagating these (both online and offline) in an effort to influence viewers, force media coverage, and provoke institutional responses.* Sometimes, manipulators choose viral slogans that attempt to co-opt an existing controversy or news topic. Other times, manipulators will attempt to fill "data voids"[8]—combinations of terms or search queries with little existing content, which can be easily associated with political messaging. Viral slogans can be spread through memes, hashtags, posters, and videos. Most importantly, because these forms are easily transmitted and copied, they can quickly spread to public forums, both online and off, and thus become far removed from the group that created them. If manipulators are able to hide the source of the slogan and create sufficient social media circulation, mainstream media sources may provide even further amplification.

Jobs Not Mobs

In October 2018, the viral slogan "Jobs Not Mobs" moved from online fringes to national attention, where it was eventually adopted by President Trump. A November 2018 investigation by *the New York Times* details how this

8 Michael Golebiewski and danah boyd, "Data Voids: Where Missing Data Can Easily Be Exploited" (Data & Society Research Institute, May 2018), https://datasociety.net/output/data-voids-where-missing-data-can-easily-be-exploited/.

phrase moved from anonymous social media users to a presidential slogan within a few weeks, highlighting the significant Twitter, Reddit, and Facebook activity generated by those involved.[9] By exploiting immigration as a partisan wedge issue and using social media, manipulators effectively mainstreamed a far-right talking point that alleges their opposition is violent and irrational.

Jobs Not Mobs meme from subreddit r_thedonald, Friday, October 12, 2018.[10]

In recent years it has become commonplace for right-wing pundits to use the term "mob" to refer to (and

9 Keith Collins and Kevin Roose, "Tracing a Meme From the Internet's Fringe to a Republican Slogan," *The New York Times*, November 4, 2018, sec. Technology, https://www.nytimes.com/interactive/2018/11/04/technology/jobs-not-mobs.html, https://www.nytimes.com/interactive/2018/11/04/technology/jobs-not-mobs.html.

10 https://www.reddit.com/r/The_Donald/comments/9nq3kp/jobs_not_mobs/.

discredit) their political opponents: anti-fascist and other left-leaning protests, as well as groups of refugees or immigrants.[11] This rhetorical strategy occurs both in fringe social media circles and mainstream media outlets such as Fox News. In 2018, Twitter users began using "mobs" in configurations of a viral slogan: "Jobs Not Mobs."

While social media posts iterating jobs and mobs organically caught on among right and far-right influencers, the phrase experienced a jump in popularity after a group organized on Reddit took special steps to craft new media disseminating and formalizing the slogan. On Reddit, users workshopped memes to spread on social media. Video clips were created, showing decontextualized turmoil at rallies and supposed migrant caravans, serving as visual reinforcement of the "mobs" claim. Easily shareable audiovisual material, alongside the deployment of a hashtag, created opportunities for a swarm of participation, and the slogan quickly grew past its point of origin in far-right online hubs like Reddit's the_donald. While the campaign gained organic traction among nationalist communities in the US, significant bot activity has been identified in the spread of the slogan on Twitter. [12]

The slogan circulated on several small right-wing news platforms, drawing support from major right-wing social media influencers. Then, in a tweet on October 21, 2018, president Trump himself used the hashtag "#JOBSNOTMOBS!" along with an embedded video taken, without attribution, from YouTube.[13] Popular press quickly reacted to Trump's adaptation of the slogan.

11 Dan Gainor, "Media Pretend Left-Wing Mobs of Protesters Aren't Really Mobs," Text.Article, Fox News, October 13, 2018, https://www.foxnews.com/opinion/media-pretend-left-wing-mobs-of-protesters-arent-really-mobs.

12 See: https://botsentinel.com/tweet-archive?s=jobsnotmobs.

13 Sophie Weiner, "Trump's 'Jobs Not Mobs' Video Was Taken From an Anonymous GOP Fan," Splinter, November 2, 2018, https://splinternews.com/that-wild-jobs-not-mobs-video-trump-tweeted-was-taken-f-1830171490.

A subsequent Fox News segment entitled "'Jobs not mobs' Trump unveils new midterm message," effectively marking this slogan as an official statement, exposing it to mainstream TV audiences and reinforcing its effectiveness among the communities that helped spread it online.[14]

It's OK To Be White[15]

"It's Okay to Be White" (IOTBW) was a viral slogan designed to capture the narrative around contemporary representations of white supremacy and identity, adopted and popularized by a variety of reactionary communities. Drawing on the popularity of the Black Lives Matter movement, this campaign sought to use racism as a wedge issue to polarize media attention to white identity politics. The first instructions for the campaign were posted on October 24, 2017, to 4chan's right wing /pol/ board by an anonymous author, loosely outlining a plan to place simple black-and-white flyers with the titular phrase on college campuses.[16] The campaign instructions presumed an audience of high school and university students. Some of the manipulators involved set their media strategy in instructional images (see below). These images read like something out of a style guide, specifying that the flyers must not contain any additional advertisement for white supremacist groups, websites, or communities, thus obscuring their place of origin and attributing no authorship. These posts proliferated on 4chan before moving on to other forums and social media. The simple instructions caught on over the next few days, and additional flyer campaigns were largely coordinated using 4chan and Discord, a messaging app originally designed for gaming

14 "'Jobs Not Mobs': Trump Unveils New Midterm Message," Fox News, accessed May 21, 2019, http://video.foxnews.com/v/5852648609001/.

15 This case study was co-authored by Becca Lewis.

16 https://archive.4plebs.org/pol/thread/146524824/#146542738.

IOTBW instructions spread across 4chan, Stormfront, and Reddit. [18]

As individuals began hanging up actual IOTBW posters, a growing number of posts appeared on forums with images of the posters hung on college campuses and other public spaces, serving as motivational encouragement for other participants in the campaign. Waves of flyers in public spaces were deployed by an unknown number of individuals during times of high youth activity in late 2017 and 2018, like Halloween[19] or the beginning of the school year.[20]

17 April Glaser, "White Supremacists Still Have a Safe Space Online. It's Discord.," Slate Magazine, October 9, 2018, https://slate.com/technology/2018/10/discord-safe-space-white-supremacists.html.

18 https://archive.4plebs.org/pol/thread/148236228/.

19 Janell Ross, "'It's Okay to Be White' Signs and Stickers Appear on Campuses and Streets across the Country," *Washington Post*, November 3, 2017, sec. *Post Nation*, https://www.washingtonpost.com/news/post-nation/wp/2017/11/03/its-okay-to-be-white-signs-and-stickers-appear-on-campuses-and-streets-across-the-country/.

20 Caitlin Byrd, "Racist Flyers Keep Appearing on South Carolina College Campuses; Experts Expect More to Come," *Post and Courier*, February 25, 2018, https://www.postandcourier.com/politics/racist-flyers-keep-appearing-on-south-carolina-college-campuses-experts/article_3188f0ee-1742-11e8-88b5-e35653c04c82.html.

The anonymous placement of flyers in schools and universities was accompanied by simple social media posts spreading the same message, often using hashtags like #ItsOkayToBeWhite and #IOTBW. In many cases, the flyers had their intended effect—members of the public reacted negatively to the signs, decrying them as racist. This reaction led to a number of institutional responses on college campuses and, eventually, news coverage.[21] This coverage began at low-level student press and small regional journalism outlets. The participants of IOTBW then collected, archived, and amplified this low-level press, lauding it as further evidence of the campaign's success. Online influencers stepped into the campaign and monetized it, such as Milo Yiannopoulos' selling IOTBW shirts on his college campus tour. This preceded the slogan's jump to the mainstream press.

The mainstream coverage of IOTBW was only possible because of the steps taken by participants to obscure the explicitly white supremacist character of the campaign's origin. The source was often identified as those who took pictures of the flyers in public space. More mainstream far-right and reactionary communities were able to rally around the statement as a counter-point to "Black Lives Matter." At the onset of the campaign in October 2017, one 4chan poster stated in an instructional thread "Based on past media response to similar messaging, we expect the anti-white media to produce a shit-storm about these racist, hateful, bigoted fliers… with a completely innocuous message."[22] And indeed, as coverage of the posters grew, these manipulators continued to call for public attention to white identity politics, or as they referred to it "anti-white racism."

21 Taryn Finley, "'It's Okay To Be White' Signs Appear In Schools, Cities Across The U.S.," *HuffPost*, November 7, 2017, https://www.huffpost.com/entry/its-okay-to-be-white-signs_n_5a01c91ce4b0368a4e87165e.

22 https://archive.4plebs.org/pol/thread/146642618/#146643305.

Groups of IOBTW manipulators communicated, at least to each other, that the split of critical and sympathetic mainstream press coverage had always been a goal of the campaign.[23] Press coverage of the posters moved from alternate influencers on YouTube and local newspapers to larger outlets such as *The Washington Post*,[24] *The Boston Globe*,[25] and *The Daily Caller*.[26] Each of those three outlets covered both the existence of the posters in physical spaces, as well as the source of the campaign. *The Daily Caller* referenced an online forum, while both *The Boston Globe* and *The Washington Post* specifically identified 4chan. In contrast to this coverage, one Fox News commentator picked up the story but adopted the preferred frame of the manipulators, focusing on the "liberal outrage" against IOTBW posters during a broadcast on November 3, 2017. Crucially, though, this particular Fox News broadcast never made the link to 4chan or the white supremacist rhetoric involved in the campaign's origin. On 4chan itself, as the segment aired, anonymous users posted all-caps utterances like "HES SAYING IT, IT"S OKAY TO BE WHITE, HES DOING IT" and "OH SHIT HE'S GONNA COVER IOTBW."[27] The media coverage even had legislative consequences: the Australian parliament was also forced to address the slogan in October 2018, when a member proposed the

23 https://archive.4plebs.org/pol/thread/152182782/#152182782.

24 Ross, "'It's Okay to be White.'"

25 Carrie Blazina and Alyssa Meyers, "Stickers Saying 'It's Okay to Be White' Posted in Cambridge," BostonGlobe.com, November 1, 2017, https://www.bostonglobe.com/metro/2017/11/01/stickers-saying-okay-white-posted-cambridge/lQrDmpt2zisM4Ka5nbOgHK/story.html.

26 Michael Bastasch, "'It's Okay To Be White' Stickers Posted Around College Town," The Daily Caller, November 1, 2017, https://dailycaller.com/2017/11/01/its-okay-to-be-white-stickers-posted-around-college-town/.

27 https://archive.4plebs.org/pol/thread/147840097/#147841619.

slogan in a motion to condemn "anti-white racism."[28] By the measures set out in the initial directions, IOTBW was a huge success for the manipulators.

Conclusion

The consistent use of viral sloganeering, phrases that move from anonymous posts online to mainstream talking points, are expected from manipulators using small news organizations and social media influencers as opportunities to frame new media spectacles. #IOTBW and #JobsNotMobs were in essence bottom-up anonymous operations, with public figures willfully or ignorantly obscuring the materials' origins in fringe online communities. By developing the attention around these slogans in authorless, anonymous, and informal settings, manipulators can pre-empt the risk that public figures might take on by supporting the slogans developed explicitly by white supremacists. Pundits, newscasters, and politicians addressed these slogans well after they'd been laundered by more public figures that report on, and participate in, conspiracies and manipulation campaigns.[29] For example, some social media influencers are popular because they both spread conspiracy theories and then report on the uptake and refutation of those hoaxes, such is the case with pundits like Cassandra Fairbanks, Milo Yiannopoulos, and Mike

28 Alan Pyke, "Australia Comes within 2 Votes of Endorsing 4chan-Coined White Victimhood Meme as National Policy," October 15, 2018, https://thinkprogress.org/australia-ok-to-be-white-resolution-senate-4chan-900a4bb7f92a/.

29 Tyler Bridges, "'Alt-Lite' Bloggers and the Conservative Ecosystem," February 20, 2018, https://shorensteincenter.org/alt-lite-bloggers-conservative-ecosystem/; Kenneth P. Vogel, Scott Shane, and Patrick Kingsley, "How Vilification of George Soros Moved From the Fringes to the Mainstream," *The New York Times*, November 2, 2018, sec. U.S., https://www.nytimes.com/2018/10/31/us/politics/george-soros-bombs-trump.html.

Cernovich.[30] And yet, once these slogans have been widely circulated and adopted, that dilution into the general social media ecosystem masks their original, manipulative intent.

Recommendations

Viral slogans often depend on specific influencers to be amplified, especially in instances where manipulators are seeding far-right, conspiratorial, and reactionary content. It is incumbent upon journalists and platforms to understand how these viral slogans rise in attention to determine if the content's spread is organic or operational. Journalists must also understand their role in an amplification network and look out for instances where they may unwittingly call attention to a slogan that is popular only within a particular, already highly polarized community online.

30 Benkler, Yochai, Robert Faris, and Hal Roberts. 2018. *Network Propaganda: Manipulation, Disinformation, and Radicalization in American Politics.* Oxford University Press.

2. LEAK FORGERY CASE STUDIES:

- The Waters Leak
- The Macron Leak

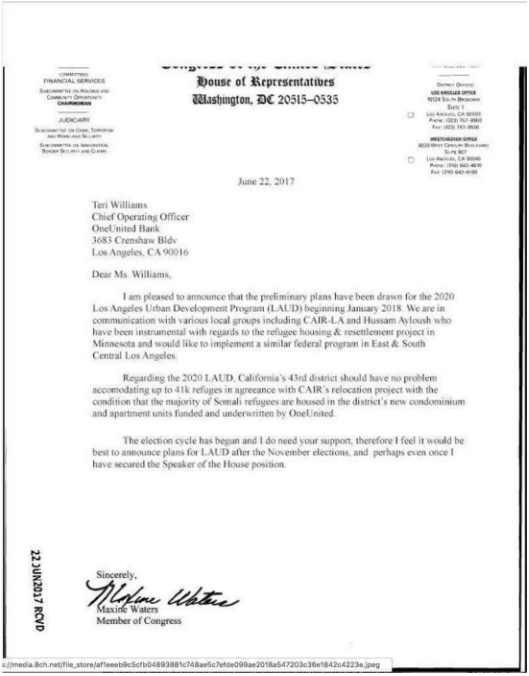
Leak Forgery *is a process of forging documents that are then released by manipulators as apparent leaks from their political targets.* As Biella Coleman has shown in her work on the "public interest hack,"[31] actual leakers can gain press exposure and political impact by revealing large packages of documents concealed from the public. Notable leaks such as those from Edward Snowden, Chelsea Manning, and WikiLeaks have set the stage for leaks to be readily accepted as a legitimate form of protest in the public interest. And because leaks often come from anonymous sources in large troves of documents, they are now fertile ground for the insertion of forged or falsified content.

Forged leaks are crafted to *appear* to indicate damaging information about political targets, who are often forced to publicly disprove the claims. Like viral slogans, forged leaks can be deployed across social media, with manipulators attempting to drum up enough apparent activity to trigger further news coverage. By staging conversation about the forged leak through alternative news outlets and social media, media manipulators draw in mainstream news coverage before any entity can debunk the documents. *Reporting on these forgeries is potentially damaging to targets, even if the leaks are later proven to be false, especially during highly contested elections.*

31 E. Gabriella Coleman, "The Public Interest Hack," Limn, May 9, 2017, https://limn.it/articles/the-public-interest-hack/.

The Waters Leak

On December 11, 2017, Republican congressional candidate Omar Navarro used Twitter to release a forgery targeting his opponent, sitting congresswoman Maxine Waters. This forged leak (see below) appeared to detail an elaborate plan by Waters to garner a donation from OneUnited Bank in exchange for allowing 41,000 immigrants to move to her district. In Southern California, where Waters has been a representative since the early 1990s, immigration has been a wedge issue in every campaign.



House of Representatives
Washington, DC 20515–0535

June 22, 2017

Teri Williams
Chief Operating Officer
OneUnited Bank
3683 Crenshaw Bldv
Los Angeles, CA 90016

Dear Ms. Williams,

I am pleased to announce that the preliminary plans have been drawn for the 2020 Los Angeles Urban Development Program (LAUD) beginning January 2018. We are in communication with various local groups including CAIR-LA and Hussam Ayloush who have been instrumental with regards to the refugee housing & resettlement project in Minnesota and would like to implement a similar federal program in East & South Central Los Angeles.

Regarding the 2020 LAUD, California's 43rd district should have no problem accomodating up to 41k refuges in agreeance with CAIR's relocation project with the condition that the majority of Somali refugees are housed in the district's new condominium and apartment units funded and underwritten by OneUnited.

The election cycle has begun and I do need your support, therefore I feel it would be best to announce plans for LAUD after the November elections, and perhaps even once I have secured the Speaker of the House position.

22JUN2017 RCVD

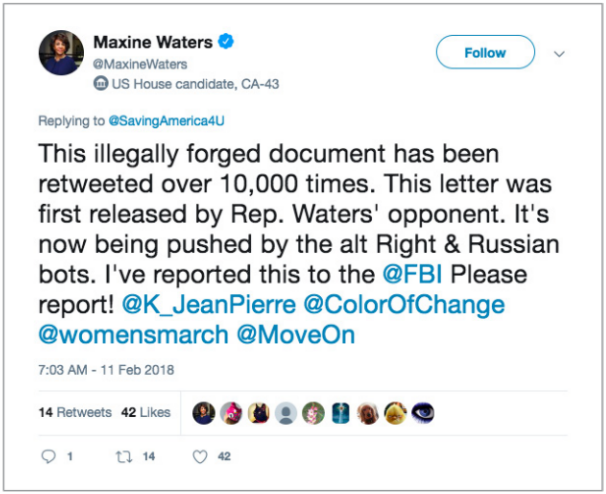
Sincerely,
Maxine Waters
Member of Congress

Screenshot of Water's Leak forgery December 11, 2017.

Over the previous year, Navarro had repeatedly and aggressively attempted to attract Maxine Waters' attention, from attacks on social media to live-streaming public protests at Waters' home. However, this forged leak did not initially get much attention on Navarro's own account. There was a spike of new attention, however, when Waters reacted by publicly requesting Twitter remove the document. *The LA Times* covered Waters' subsequent tweets calling for a Justice Department investigation into the matter. This flurry of attention allowed Navarro to claim a small victory in a YouTube video tilted "Merry Christmas — We Got Maxine Waters Attention."[32]

32 Omar Navarro, *Merry Christmas — We Got Maxine Waters Attention*, 2017, https://www.youtube.com/watch?v=5XPioTmU9yY.

The forgery gained additional traction when it was picked up on January 29, 2018, by Twitter user @SavingAmerica4U, a Twitter account known for sharing nationalist propaganda and hyperpartisan right-wing news.[33] @SavingAmerica4u's tweet received significant interactions and was then spread on Reddit, particularly on r/the_donald. Attention was drawn to Navarro and his campaign through both organic interactions with the material and artificial amplification by bots and sockpuppets, creating a swarm that helped drown out the claims of forgery. As of October 2018, the tweet containing the forgery had nearly 15,000 retweets, 12,000 likes, and remained online.



Maxine Waters drawing attention to the forged leak on Twitter, February 11, 2018.

33 Saving America, "Well, Well, Well... BUSTED Maxine Waters Tells United Bank That She Needs Their Money to Get Elected but Doesn't Want Anyone to Mention the 41,000 Somalian Refugees Being Re-Located in LA until after the November Elections Bc It May Hurt Her Chances of Re-Election.Pic.Twitter.Com/Nd6GOSy9nJ," Tweet, *@SavingAmerica4U* (blog), January 29, 2018, https://twitter.com/SavingAmerica4U/status/958016521112031232.

The image of the forged leak was not accompanied by any source in Navarro's or @SavingAmerica4u's tweets. A stray bit of metadata, a URL embedded in the bottom of the screenshotted document cropped out of the Twitter preview thumbnail, unintentionally revealed the original source of the image. This URL links back to 8chan, an anonymous image board which, like 4chan, is an arena for the development of conspiracy theories and other false leaks. 8chan is unindexed, not searchable by Google or other conventional means, difficult to navigate, and rife with extreme racist propaganda and all manner of highly offensive and illegal content. The Waters' leak also appears suspiciously similar to the distribution of forged documents during the French election in 2017, which was also linked back to a series of posts on 4chan.

The Macron Leak

On Wednesday, May 3, 2017, an anonymous 4chan user posted a series of links and instructions titled *Documents proving Macron's secret tax evasion*.[34] The links were to nine gigabytes of files allegedly obtained via a phishing scam from the digital archives of French presidential candidate Emmanuel Macron. As comments on these posts grew, so did a call to swarm the documents, as one user on 4chan posted, "See what you can do with this, anon. Let's get grinding. If we can get #MacronCacheCash trending in France for the debates tonight, it might discourage French voters from voting Macron." As the campaign spread across multiple threads on 4chan and other message boards, several unverified documents indicating tax evasion and offshore account holdings surfaced in replies.

34 https://archive.4plebs.org/pol/thread/123933076/.

Fax

TO: Richard Palmer
COMPANY: La Providence Ltd.
TEL: 1-888-469-2829
FAX NO.: 1-888-469-6509

FROM: Brian Hydes
TEL: 1-345-815-2252

DATE: 11/02/15
NO. OF PAGES: 2
COPY:

CIBC FirstCaribbean International Bank

CIBC FirstCarribean (Cayman) Limited
PO BOX 66 GT
25 Main Street
Grand Cayman, Cayman Islands

www.firstcaribbeanbank.com
Tel: (345) 949-7300
Fax: (345) 949-7179

RE: Returned Cheque

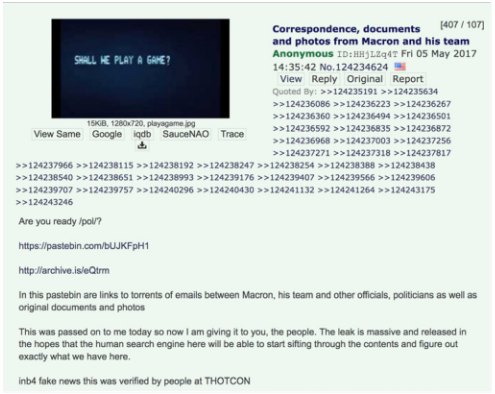
Dear Mr. Palmer,

Please find the attached copy of the returned cheque as requested. The original was sent via mail.

Regards,

Brian Hydes
Transactional Banking Associate
International Banking Centre

A forged document claiming Macron had an overseas bank account.



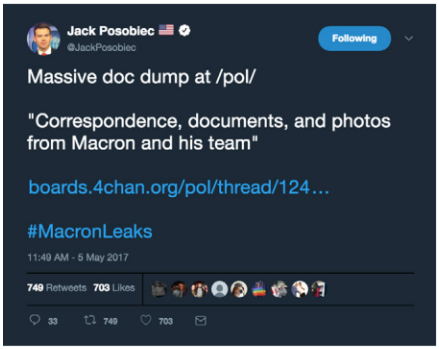
Post announcing the location of the Macron Leak.

On 4chan, users began a campaign to spread the leaked documents on social media, in an attempt to damage Macron before his race against far-right opponent Marine Le Pen. "Send it to non French journalists, spam it on Twitter,"[35] posted one user. French law dictates a 44-hour press blackout leading up to the election, and thus major domestic media outlets were legally unable to comment on the manipulation campaign as it spread. In the period of press silence on the campaign, these

35 https://archive.4plebs.org/pol/thread/123933076/#123933592.

leaks were amplified by social media users and political operatives, particularly WikiLeaks. On May 5, Macron's campaign was forced to formally acknowledge[36] the phishing scam that led to the spread of real and falsified documents circulating online.

This thread singled out several documents alleging to find Macron's dirty cash.



An American conspiracy theorist posts a link to the investigation on 4chan.

36 Christopher Dickey, "Did Macron Outsmart Campaign Hackers?" *Daily Beast*, May 6, 2017, https://www.thedailybeast.com/did-macron-outsmart-campaign-hackers.

By quickly replicating and disseminating real and forged content and taking advantage of the limitations of mainstream media's ability to verify this information, 4chan users saturated the online conversation with a mix of forged leaks and adversarial political memes.[37] In addition, alternative news reporting helped populate search engines and social media with content critical of Macron, obscuring the source of this multi-operator campaign. While the campaign was thoroughly debunked[38] in mainstream media, the impact of these forged leaks was considerable—actors behind a 4chan manipulation campaign were able to capture the attention of social media users and the press.

The nesting of fake content within a larger dataset of leaked material complicates verification and debunking efforts. As detailed in their 2017 report *Tainted Leaks Disinformation and Phishing With a Russian Nexus*, Citizen Lab examines how such "tainted" information dumps online attract media and civic attention, that inadvertently amplifies "enticing, but questionable information" alongside legitimate documents.[39] The effect of this on the public's ability to assess the veracity of leaks in the public interest is called into question by this new tactic.

37 Perrine Signoret, "Sur 4Chan, les anti-Macron diffusent mèmes accusateurs et fausses rumeurs," l'express, April 24, 2017, https://lexpansion.lexpress.fr/high-tech/sur-4chan-les-anti-macron-diffusent-memes-accusateurs-et-fausses-rumeurs_1901827.html.

38 "How We Debunked Rumours That Macron Has an Offshore Account," The Observers, May 5, 2017, https://observers.france24.com/en/20170505-france-elections-macron-lepen-offshore-bahamas-debunked.

39 Adam Hulcoop et al., "Tainted Leaks: Disinformation and Phishing With a Russian Nexus," The Citizen Lab, May 25, 2017, https://citizenlab.ca/2017/05/tainted-leaks-disinformation-phish/.

Conclusion

Forged leaks are most effective during their initial circulation, before the documents can be subjected to scrutiny and authenticity checks. Even after fact-checking, the damage may already be done by negative or erroneous coverage. Coverage of leaked forgeries allows manipulators to put their target on the defensive. Manipulators in these cases may be motivated by political agendas, though there are many reasons for the participation in the spreading of fake leaks online. Leakers may be generating traffic for advertiser-driven websites reporting on these forgeries, and thus share for profit. In either case, the supposed status of "leaked" forgeries as private or secret documents is an effective cover for those manipulators creating the false content.

Recommendations

Reporters and platform technologists must check the provenance of leaked materials and verify the source as well as the target. If there is no attribution and the hoax involves multiple institutional actors, seek numerous confirmations from unrelated sources. It is a red flag if there is a clear wedge issue at play, like in the Waters' case, where platforms should take special care to not allow this content to spread. Journalists should also coordinate fact-checking efforts, where possible, especially in high-stakes coverage of elections and other events of national or international importance.

3. EVIDENCE COLLAGES CASE STUDIES:

- Charlottesville Unite the Right Rally
- Pizzagate

Evidence collages *are image files featuring a series of screenshots and text that arrange evidence of a particular event or activity.* Manipulators use these carefully constructed "infographics" to sway breaking reporting and encourage further investigation by citizens. Evidence collages often contain a mix of verified and unverified information and can be created with simple image-editing software. By compiling a mix of verified and unverified information in a single, shareable image, conspiracy theorists can decide which sources to highlight or obscure, bypassing critical search results a user would experience if they were to seek this information on their own. Graphic design choices lead audiences through guided pathways of information, using keywords as dog whistles intermingled with URLs containing further disinformation. By mimicking the presentation of an infographic or official document, manipulators can use evidence collages to guide viewers to investigate and join the campaign by supporting sites, resources, hashtags, and communities.

The creation of manipulative evidence collages has developed alongside the growing prevalence of open source investigation (OSI) techniques. OSI techniques use available search engines and databases to gather information based on small pieces of evidence in live breaking photos or video. These techniques are now common responses to breaking news events, especially instances of high-profile violent crime. Police may turn to social media and other

sources for traces of suspects' activities. Reporters might dig through social media for publishable facts. Everyday users may even see themselves as participants in a form of investigation, such as with the failed investigation attempts conducted at Reddit in the wake of the Boston Marathon bombing of 2013.[40]

Manipulation campaigns can use these same techniques, combing through various databases of media and other information, to create new politically targeted visual narratives. Streams hosted on YouTube and Periscope can be combed for evidence and identifiers, which are then compiled at discussion sites, like 4chan or Reddit. Manipulators can use personal information indexes to look up suspected actors' names, addresses, and vehicle registrations in an effort to ascertain their possible motivations. These bits of evidence are then shared, discussed, and compiled in threads and link repositories. However, unlike official investigators, manipulators can pick, choose, or falsify the media they compile to steer their own political narrative.

40 Alex Leavitt, "Upvoting the News: Breaking News Aggregation, Crowd Collaboration, and Algorithm-Driven Attention on Reddit.Com" (University of Southern California, 2016).

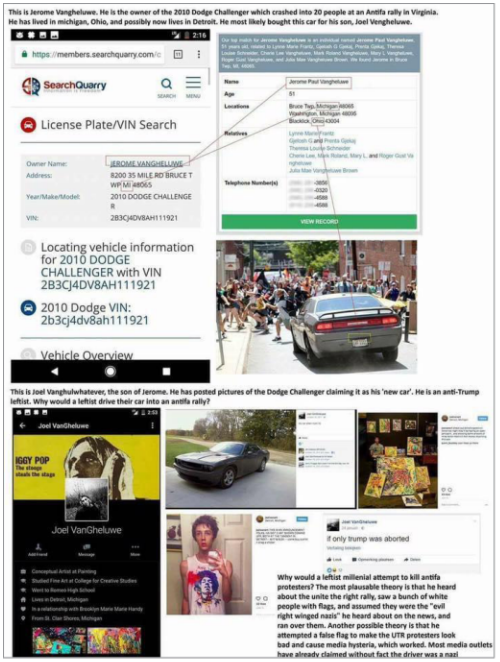
Charlottesville Unite the Right Rally[41]

The Unite the Right Rally on August 12, 2017, in Charlottesville, Virginia, was the largest white supremacist rally in modern US history. Months prior to the rally, local residents pleaded with city officials to prevent it from happening, citing several other violent rallies earlier that year led by the newly formed "alt-right" movement and the KKK. The politicization of the Robert E. Lee monument began a year earlier when local activists persuaded city officials to remove it from a public park. White supremacists seized the moment to gain media attention to their cause by claiming the removal of the statue was "anti-white."

During the August 12 rally, James Fields, a white supremacist, drove his car into a group of anti-racist counter-protestors, injuring several and killing Heather Heyer. After the fatal impact, Fields reversed and fled the scene.[42] In the time before the suspect was apprehended and identified, the public attention around the attack provided an opportunity to manipulate public conversation. Using open source investigation techniques, 4chan users constructed a convincing evidence collage, falsely identifying the driver of the car, which was then amplified on far-right news platforms.

41 This case study was co-authored with Peter Krafft, a Data & Society fellow, 2017.

42 Joe Heim, "Recounting a Day of Rage, Hate, Violence and Death," *Washington Post*, August 14, 2017, https://www.washingtonpost.com/graphics/2017/local/charlottesville-timeline/.



Evidence Collage taken from "PuppetString News" Website.[43]

In the wake of the attack, hundreds of posts posited theories of the crime with a groundswell of screenshots to support various findings. In just a matter of hours, several collages (see above) appeared on 4chan that presented evidence that the Dodge Challenger involved in the Charlottesville rally was driven by a young left-leaning student named "Joel V." Central to these collages was a high-resolution photo of the back of the car that included the car's license plate. A 4chan/pol/ user posted that they "ran the plates" on a website called "Search Quarry" and then provided visual evidence of the driver registration associated with the license plate. It had previously been registered to a man named Jerome

43 Original URL: https://www.puppetstringnews.com/blog/friendly-firecharlottesville-car-attacker-is-anti-trump-antifa-supporter.

V. Moreover, Joel's surname is so rare online that search engines instantly returned his personal accounts. Manipulators then took social media posts from Jerome's son Joel to create a disinformation narrative that framed Joel as a disgruntled leftist.

While such evidence collages might give the impression of a legitimate group investigation, careful attention to details of the data craft can often reveal that these collages are deliberately designed by single actors attempting to reframe narratives in their favor. These authorship clues involve examining the collage(s) for visible metadata, such as battery life, open tabs or apps, and time stamps. In the evidence collage that frames Joel V., two components reveal these context clues, including the screenshots of "Search Quarry" and his Facebook page above the Iggy Pop banner. The similarities in battery life, time stamp, and open apps indicate that these two pieces of evidence were taken from the same phone.

Nevertheless, once false pieces of information are packaged in these types of evidence collages, the images can be easily moved from anonymous message boards to mainstream news and information sources. Social media is designed to allow easy sharing and amplification of images, where provenance is easily lost, if the person sharing does not take care to retain this data.[44] Such collections of "evidence" can also be the catalyst for alternative news coverage on blogs, creating a publishing track record that can be cited by larger and more legitimate news sites.[45]

44 Laura Mallonee, "How Photos Fuel the Spread of Fake News," *Wired*, December 21, 2016, https://www.wired.com/2016/12/photos-fuel-spread-fake-news/.

45 See: https://pastebin.com/dpWqVTwj.

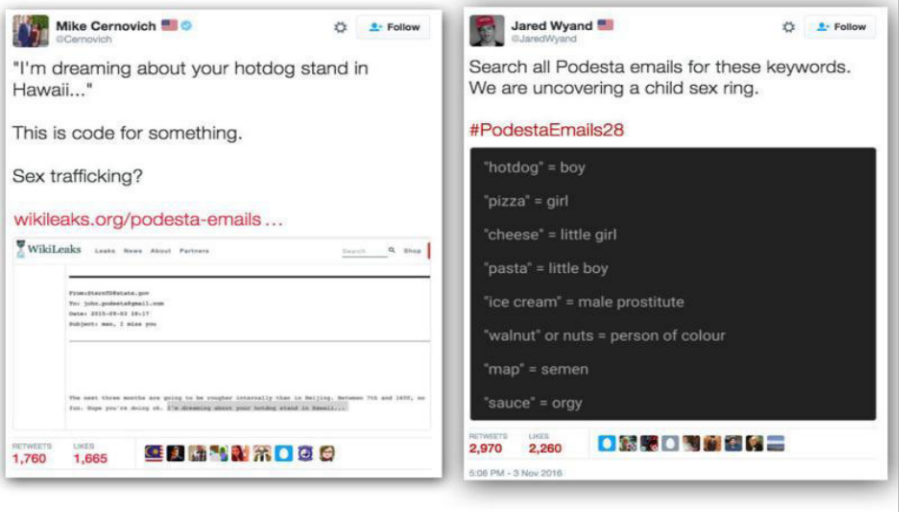
In the case of the Charlottesville rally, numerous alt-right accounts, blogs, and message boards used these spurious evidence collages to shape media narratives. While eventually debunked, the claims that "Antifa" was responsible for the attack gained some initial traction during the feverish news cycle immediately following the event. Conservative media platforms, such as, Gateway Pundit, GotNews, and Puppet String News, all shared the story, often embedding a cropped version of the 4chan-sourced collages. The narrative and images eventually also appeared on FreeRepublic, Zerohedge, Reddit, and Twitter. The collage was posted in the replies of many "breaking news" tweets from mainstream media outlets, where it sparked conversation and encouraged harassment of the misidentified. The consequences for Joel and his family were high, as they had to flee their home and hide for several days.[46]

Pizzagate

Pizzagate was a popular online conspiracy begun in 2016 that posited prominent Democrats and world leaders were engaged in clandestine secret societies trafficking in pedophilia. Emerging out of a Wikileaks document dump of emails from political consultant John Podesta in 2016 and fueled by a decade of amateur investigations into Jeffrey Epstein, creators of Pizzagate's evidence collages synthesized cultural paranoia regarding sex trafficking to guide audiences into an alternative information sphere. In this series of visual narratives, evidence proliferated that Podesta and Hillary Clinton were using "pizza" as code for ordering sex acts with children. Manipulators provided elaborate ways to "decode" the language in the

46 Avery Anapol, "Man Misidentified as Charlottesville Driver Sues Far-Right Websites for Defamation," TheHill, February 17, 2018, https://thehill.com/blogs/blog-briefing-room/374412-man-misidentified-as-charlottesville-driver-sues-far-right-websites.

documents in the Podesta leaks (see below). These codes, and other elaborate evidence collages were spread on forums and social media, designed to be compared, discussed, and used as the beginning of a guided research process.



Tweets by far-right pundits pushing the theory that food is code for child sex trafficking.[47]

47 Craig Silverman, "How the Bizarre Conspiracy Theory Behind 'Pizzagate' Was Spread," *Buzzfeed News*, November 4, 2016, https://www.buzzfeed.com/craigsilverman/fever-swamp-election?utm_term=.uqlpalZrL#.phq9LeqVK.

Pizzagate Infographic #1, November 2016.[48]

In one series of Pizzagate Evidence collages (see above), the designer turns from using alternative archive links and screenshots to more unverifiable sources. Increasingly obscene, disturbing, or leading imagery is juxtaposed against doctored images of mainstream press and reproductions of 4chan threads. The use of evidence collaging by manipulators created a powerful, disturbing closed environment where the creator commanded attention, authority, and authenticity.

Pizzagate was not the product of a single campaign organized around a breaking news event. Rather, it was an extended effort across multiple platforms. Some spreading the meme sought to expose politicians for the scandal, while others attempted to equate pedophilia and homosexuality. While LGBTQ acceptance has been a mainstay of political discourse over the last two decades, virulent homophobia and transphobia online are significant

48 https://imgur.com/gallery/QgS7z.

wedge issues that can be memetically mobilized. Those pushing Pizzagate out of the shadows of message boards and into more reputable outlets relied on initial coverage from niche blogs, amplification on alternative news sites, and the social media presence of influential conspiracy theorists to make it go viral.[49] Bolstered by an alternative news system of conspiracy theorists, pro-Pizzagate coverage dominated alternative news, conspiracy outlets, and independent blogs before fact-checked news sources provided critical reporting on the campaign.

Rolling Stone investigated the spread of Pizzagate across Twitter and found that bots and right-wing pundits took hold of the story slowly, but as election day approached in 2016, the attention to the conspiracy swelled as blogs and message boards provided detailed overviews of the complex conspiracy.[50] The use of widely distributed evidence collages, analysis of this evidence on blogs and forums, and resulting explainer videos on YouTube propelled this story to new audiences. During December 2016, as the activity around Pizzagate moved from an examination of leaks critical to the Democratic National Convention to calls to action against "satanic" ritualistic pedophile rings, major platforms like Reddit sought to ban Pizzagate content, forcing manipulators to seek other platforms to organize and workshop materials.

These calls to action led several believers of the conspiracy to visit Comet Ping Pong, a DC pizza restaurant referenced in the Podesta email dump as an organizing hub for a global pedophile ring. On December 6, 2017, a gunman named Edgar Maddison Welch entered Comet with an armed rifle and demanded he be allowed to investigate. He fired several rounds before surrendering

49 Benkler, Faris, and Roberts, *Network Propaganda*, 226–230.

50 Amanda Robb, "Pizzagate: Anatomy of a Fake News Scandal," *Rolling Stone*, November 16, 2017, https://www.rollingstone.com/politics/news/pizzagate-anatomy-of-a-fake-news-scandal-w511904.

to police.[51] Welch was later sentenced to four years in prison, and prominent conspiracy theorists like Alex Jones subsequently distanced themselves from the campaign.[52] While the campaign has faded from popularity, the materials it produced, primarily the evidence collages, still exist online, and help inform subsequent conspiracy theories like Qanon.[53]

Conclusion

Evidence collages are a recurring tactic used by manipulators during breaking events and in support of harassment campaigns or conspiracy theories. Evidence collages can degrade the public trust in established news outlets, public accountability, and the veracity of public conversations themselves during times of ongoing investigations and media spectacles. Like forged leaks, evidence collages often combine verifiable and unverifiable information. Evidence collages are an alternative framing device that helps hide the source of disinformation, particularly when links and archives to blogs and other websites are included.

51 Matthew Haag and Maya Salam, "Gunman in 'Pizzagate' Shooting Is Sentenced to 4 Years in Prison," *The New York Times*, January 20, 2018, sec. U.S., https://www.nytimes.com/2017/06/22/us/pizzagate-attack-sentence.html.

52 James Doubek, "Conspiracy Theorist Alex Jones Apologizes For Promoting 'Pizzagate,'" NPR.org, March 26, 2017, https://www.npr.org/sections/thetwo-way/2017/03/26/521545788/conspiracy-theorist-alex-jones-apologizes-for-promoting-pizzagate.

53 Renee Diresta, "Online Conspiracy Groups Are A Lot Like Cults," November 13, 2018, https://www.wired.com/story/online-conspiracy-groups-qanon-cults/.

Recommendations

Do not repost collaged materials without vetting and confirming the source. To spot this kind of campaign, journalists and platform companies should seek out signs of cross-platform coordination, especially for image files that circulate without clear authorship. Because manipulators will use different methods to hide the origins of these images, use reverse image software to look for other possible sites. Use timestamps and the stray metadata contained within screenshots to infer if this is the work of an individual or group. If the identity of an individual is involved, do not recirculate without outside confirmation from a well-regarded source. Because these images are intended to sway public conversation, they usually end up in the replies of social media feeds of news organizations. Where possible, social media moderators should take care to downgrade or hide evidence collages containing misinformation, if the platform allows this. Platforms acting as communication infrastructure companies should enhance the capacity for account holders and moderators to control content and comments on their pages.

4. KEYWORD SQUATTING CASE STUDIES:

- Antifa Social Media Accounts
- Internet Research Agency

Keyword squatting is a technique of creating social media accounts or content associated with specific terms to capture and control future search traffic. This technique of "squatting," takes its name from "domain squatting," where individuals register web domains they believe will later become valuable, in the hopes of turning a profit.[54] These types of squatting can also support forms of online impersonation, where manipulators use misleading account names, URLs, or keywords to speak as their opponents or targets. Manipulators can use search engine optimization strategies to leverage hashtags and content tags to surface their wares above genuine sources of information. While marketers have used such strategies in the past to gain attention for brands, here we see the strategic co-opting of keywords related to breaking news events, social movements, celebrities, and wedge issues. Using keywords, manipulators will overpopulate tagged conversations using imposter accounts with fabricated attribution and bad-faith commentary. In some manipulation campaigns, they preemptively claim usernames and hashtags for strategic impersonation or parody, which become effective tools for controlling opposition or scapegoating activists or minoritized groups.

54 Janos Szurdi and Nicolas Christin, "Domain Registration Policy Strategies and the Fight against Online Crime," n.d., 16; "Username Squatting Policy," accessed May 23, 2019, https://help.twitter.com/en/rules-and-policies/twitter-username-squatting.

Antifa Social Media Accounts[55]

Keyword squatting has been most effective when used to impersonate loosely affiliated social movement organizations. A notable instance was the proliferation of fake Antifa accounts created in 2017. Increasingly visible at public protests around the US and Europe, Antifa are not a traditional political organization, and as such, do not maintain a verified presence on social media. Various white supremacist groups have consistently tried to damage Antifa's reputation in the media by "doxing" protesters (releasing their personal information) or impersonating them online. Throughout 2017, right-wing manipulators utilized parody to discredit Antifa, taking advantage of available Twitter handles and public confusion about the organization and their motives.

Collage by Twitter user @RVAWonk, April 2017. [56]

55 This case study was co-authored with Matt Goerzen, a Researcher at Data & Society.

56 Caroline Orr, "Someone Has Created a Lot of Fake Antifa Twitter Accts over the Past 2 Mos. Pretty Sure I Know Who. More Info Later. But Don't Fall for It.Pic.Twitter.Com/VffoGCwDfk," Tweet, *@rvawonk* (blog), April 25, 2017, https://twitter.com/rvawonk/status/857011182988873728.

One notable fake Antifa account gained attention on February 17, 2017, when a Facebook page associated with a Boston-area chapter was removed from the site.[57] In response, two new "Boston Antifa" pages quickly emerged. These linked accounts were fake and used satire to imitate Antifa beliefs; the imposters produced content that mocked Antifa talking points, placing emphasis on perceived inconsistencies in Antifa's philosophy. The misattribution and disinformation sown by these accounts succeeded, based on anonymous organizing – both of the anonymity of the campaign's creators, as well as Antifa activists – and as such, avoided initial attempts to debunk.

Several other false Antifa accounts were discovered by investigators in 2017, attributed to a call to action posted on 4chan.[58] As revealed by Bellingcat investigator Elliot Higgins,[59] manipulators spread false claims that Antifa were calling for organized violence against female Trump supporters using the hashtag #PunchANazi. This gruesome campaign reused images from an advocacy campaign against domestic violence featuring images of battered women and children. Other fake Antifa accounts were easier to spot as parody, with ironic titles ("Official Antifa" or "Beverly Hills Antifa") and subtle references to 4chan memes. Many of these accounts baited journalists and influenced public figures by taking advantage of the decentralized nature of Antifa and the

57 Anonymous Contributor, "Alt-Right Trolls Are Posing as Boston Antifa on Facebook and YouTube," *It's Going Down* (blog), March 2, 2017, https://itsgoingdown.org/alt-right-trolls-posing-boston-antifa-facebook-youtube/.

58 "Far-Right in Smear Campaign against Antifa," August 24, 2017, sec. BBC Trending, https://www.bbc.com/news/blogs-trending-41036631.

59 Eliot Higgins, "White Nationalists on 4chan Start a Fake Social Media Campaign to Smear #Antifa as Promoting the Targeting of White Women for Violencepic.Twitter.Com/XIsHzDWg4O," Tweet, *@eliothiggins* (blog), August 23, 2017, https://twitter.com/eliothiggins/status/900606200479404032.

pseudonymous nature of its public communications. One account even received credulous, mistaken coverage in the *New York Times*.[60]

Another notable instance of fake Antifa activity came in July of 2017. Kevin Stafford, A YouTuber who made a series of "Boston Antifa" parody videos, secured an invitation to appear as a representative of Antifa on Fox News in an interview with Jesse Watters.[61] Watters believed Stafford was the author of a post on the anarchist blog ItsGoingDown.org.[62] This article, "An Open Letter to Liberals and Progressives from the Black Bloc," was a call for liberals to abandon support for the political system.[63] During the interview, Stafford played a caricature of an Antifa member, at one point claiming that police horses could be racist supporters of Donald Trump. After the broadcast, it was revealed that Stafford was a self-identified YouTube "troll" who posts under the name BG Kumbi.[64] Despite this revelation, Fox News tweeted a quote and clip from Stafford's interview (see below). Even after Stafford himself acknowledged his duplicity, Fox News' Twitter post remained, uncorrected and unqualified. Here, then, source hacking works in two ways: with Stafford hiding his own identity from Fox News and Fox News' tweet continuing to circulate to new audiences.

60 Bari Weiss, "Opinion | We're All Fascists Now," *the New York Times*, July 31, 2018, sec. Opinion, https://www.nytimes.com/2018/03/07/opinion/were-all-fascists-now.html.

61 For the full video of the interview, see: https://www.youtube.com/watch?v=vmlyePmLIVE.

62 It's Going Down, "Fox News Hosts Alt-Right Trolls to Talk About IGD Article," *It's Going Down* (blog), July 17, 2017, https://itsgoingdown.org/fox-news-hosts-alt-right-trolls-talk-igd-article/.

63 This letter can be found here: https://itsgoingdown.org/open-letter-liberals-progressives-black-bloc/.

64 Blake Montgomery, "A Fake Antifa Member Trolled A Fox News Host, And The President May Have Watched It," *BuzzFeed News*, July 18, 2017, https://www.buzzfeednews.com/article/blakemontgomery/a-fake-antifa-member-trolled-a-fox-news-host-and-the.

Fox News tweet promoting Jesse Watters' interview with Kevin Stafford, July 2017. [65]

It has been noted by other researchers that "Antifa" as a keyword is disproportionally used by accounts associated with the right and troll accounts.[66] This allows anonymous manipulators to further dominate "Antifa" social media searches. In response to the propagation of

65 Fox News, "*.@JesseBWatters*: 'What about When an ANTIFA Member Stabbed a Police Horse...Was the Horse a Racist Trump Supporter?' ANTIFA Member: 'Yes.'Pic.Twitter.Com/MtwSmG5alr," Tweet, *@FoxNews* (blog), July 16, 2017, https://twitter.com/FoxNews/status/886701219841847296.

66 Conspirador Norteño, "The Social Construction of Antifa as Terrorists by the Alt-Right," Tweet, *@conspirator0* (blog), January 9, 2018, https://twitter.com/conspirator0/status/950963660582735872.

fakes, an "Antifa checker" Twitter profile[67] was created to help verify legitimate accounts. Keyword squatting has become a standard tactic used to discredit social movements, where an initial rush of activity and public interest helps manipulators identify unpopulated terms and seed them with biased content. Moreover, when used effectively, keyword squatting allows for controlled opposition on a wedge issue. In this example, the keyword "Antifa" was easily taken over by those who sought to mischaracterize the social media presence of people who were challenging both Trump and the "alt-right" in street protests.

Internet Research Agency[68]

Foreign actors have been shown to utilize keyword squatting to exploit wedge issues in US culture. The failure of the US political system to adequately address racism and police brutality has become a national security issue because of its capacity to be weaponized on social media. Specifically, many journalists have focused on 3,000 ads purchased by 470 "suspicious and likely fraudulent" Facebook accounts and pages linked back to a Russian company called the Internet Research Agency (IRA).[69] In early September, Facebook confirmed that these pages were linked to this Russian "troll farm."[70] In early October 2017, Facebook admitted that as many as 10 million Facebook users may have seen the ads purchased by these pages. The ads are said to have focused on a number of wedge issues, not just straightforward candidate endorsements and attacks.[71]

67 See: https://twitter.com/antifachecker?lang=en.

68 This case study was co-authored by Becca Lewis.

69 "Authenticity Matters: The IRA Has No Place on Facebook | Facebook Newsroom," accessed May 23, 2019, https://newsroom.fb.com/news/2018/04/authenticity-matters/.

70 Adrian Chen, "The Agency," *The New York Times*, June 2, 2015, http://www.nytimes.com/2015/06/07/magazine/the-agency.html.

71 John Shinal, "Facebook Says 10 Million People Saw Russian-Bought Political Ads," CNBC, October 2, 2017, https://www.cnbc.com/2017/10/02/facebook-says-10-million-people-saw-russian-bought-political-ads.html.

Screenshot of IRA Facebook Page "Blacktivist" ad.[72]

Jonathan Albright, the Research Director at the Tow Center for Digital Journalism, released a dataset that showed that Russian influence on Facebook likely extended well beyond these paid ads. Specifically, the Internet Research Agency adopted an organic marketing approach of cultivating a following over time on their fraudulent pages, aided in part by directly paying for ads and promotion. The 470 pages linked back to the IRA-mimicked-activism-oriented Facebook communities and thus could draw in users interested in certain communities and issues. These topics ranged from political and religious affiliation, patriotism, Black activism, conservatism, and LGBT issues, among others. The IRA was able to attract people with highly targeted messages by making appeals to their emotional responses around

72 https://russian-ira-facebook-ads.datasettes.com/russian-ads-919cbfd/display_ads?_target=e95f6.

sociocultural, political, and moral questions using a mix of impersonation, organic reach, and digital advertising.[73]

Faked black advocacy groups gained popular traction on Facebook in 2016, specifically popular pages "Blacktivist" and "Black Matters." Activists from the Black Lives Matter movement reported these pages to Facebook, but received no response. Later revealed by CNN both were IRA-run pages and have since been removed.[74] IRA administrators on these pages posted content, replied to messages, and organized events, all without revealing their identity to audiences. Metric cues, such as hundreds or thousands of followers, made these accounts appear legitimate. Moreover, their co-presence on other social media platforms, such as Instagram and Twitter, further insinuated that these accounts were well-resourced social movement organizations. The IRA use of Facebook pages showed how tech-savvy manipulators could frame a multi-tier, multi-issue manipulation campaign across hundreds of accounts by impersonating social movement organizations through strategic use of keywords.

Conclusion

Tactically, keyword squatting relies on social media users' practices of searching for information on social and political issues. Because social media platforms do not surface content outside their owned domains, manipulators must use data craft to ensure that their content is optimized

73 Anthony Nadler, Matthew Crain, and Joan Donovan, "Weaponizing the Digital Influence Machine," *Data & Society* (blog), October 17, 2018, https://datasociety.net/output/weaponizing-the-digital-influence-machine/.

74 Donie O'Sullivan and Dylan Byers, "Exclusive: Fake Black Activist Social Media Accounts Linked to Russian Government," CNN Business, September 28, 2017, https://money.cnn.com/2017/09/28/media/blacktivist-russia-facebook-twitter/index.html.

within each platform and is also spread across multiple platforms. When manipulators impersonate social movements, they can affect cultural attitudes, politics, reporting, and even policy.

Recommendations

These particular kinds of campaigns are very difficult to spot from outside the platform. Social media companies have access to far more of the metadata that can reveal the origins of specific accounts, memes, and slogans. Facebook has begun to blog about their efforts to remove "coordinated inauthentic behavior."[75] Twitter is also providing publicly available spreadsheets of accounts used by Russian operatives.[76] These efforts should be expanded and regularly re-evaluated for reliability and quality checks.

Yet the efforts manipulators will go to in order to hide their coordination can vary from no protections to laundered banking information. Especially in instances where hoaxes involve numerous platforms, assessing the source materials circulated for coherency with other movement organizations is paramount. However, this is not foolproof, as skilled manipulators will share a mix of content from other easy-to-verify sources. If the account or page links back to specific domains or solicits donations that are not associated with any identifiable person or entity, then it should be considered a red flag. Platform companies should make this metadata more accessible on the "about" or "bio" pages of accounts.

75 "Removing Coordinated Inauthentic Behavior from Russia | Facebook Newsroom," January 17, 2019, https://newsroom.fb.com/news/2019/01/removing-cib-from-russia/.

76 "Elections Integrity," accessed May 23, 2019, https://about.twitter.com/en_us/values/elections-integrity.html.

CONCLUSION

These four tactics of source hacking work because networked communication is vulnerable to many different styles of attack, and finding proof of coordination is not easy to detect. Source hacking techniques complement each other and are often used simultaneously during active manipulation campaigns. These techniques may be carefully coordinated, but often rely on partisan support and buy-in from audiences, influencers, and journalists alike. Viral sloganeering allows small groups of manipulators to receive disproportionate mainstream coverage by encouraging those exposed to their slogans to seek further information online. Forged leaks are seeded by manipulators and set the stage to defame public figures. Similarly, the creators of evidence collages amplify falsified documents and propaganda to sway journalistic coverage and prompt audiences to self-investigate. Keyword squatting allows manipulators to impersonate individuals and organizations, creating false impressions of their targets' goals and allowing for controlled opposition.

Manipulators who use the techniques illustrated here rely on quick deployment and prior organizing experiences to coordinate participation. Manipulation campaigns that gather on one platform to plan an attack on another are designed to give the impression of large-scale public engagement. This adversarial media environment requires both journalists and platform designers to think with the tools of information security and open source intelligence to spot when they are being manipulated. Greater attention to the coordination of manipulation campaigns across platforms is the most productive way to guard against their reach. Only through careful attention to the data craft used to create disinformation can these campaigns be debunked in a timely manner.[77]

77 Acker, "Data Craft."

Likewise, best practices for the proper identification of bad actors and artificial amplification tactics on social media are needed to prevent small groups from polarizing public conversation through wedge issues. Authorship and attribution of disinformation are often impossible to ascertain, and public figures actively participating in manipulation campaigns use all media attention, positive or negative, as an opportunity for audience building.[78] For those individuals who participate in manipulation campaigns, we cannot measure intent—we can only measure harm.[79] And so, gauging the true intentions of manipulators is not the most critical issue facing those fighting against manipulation campaigns—gauging their impact is.

In this adversarial media environment, little can be done to prevent source hacking. Manipulation campaigns are a feature of contemporary social media, where advertisers, trolls, spies, hackers, grifters, scammers, public relations companies, politicians, and many others maneuver for attention to their issues. Unless platform companies, researchers, and journalists begin to pay careful attention to these tactics, campaigns will continue to evolve. As we have shown, simply paying attention to anonymous message boards could also do damage because manipulators are often conscious they are being watched and will turn it to their advantage. Social engineering, which includes technical and psychological manipulation, is built into sociotechnical systems, where journalists may be unaware that they are the targeted adversary.

78 Molly McKew, "How Liberals Amped Up a Parkland Shooting Conspiracy Theory," *Wired*, February 27, 2018, https://www.wired.com/story/how-liberals-amped-up-a-parkland-shooting-conspiracy-theory/.

79 Evelyn Douek, "Facebook's Role in the Genocide in Myanmar: New Reporting Complicates the Narrative," Lawfare, October 22, 2018, https://www.lawfareblog.com/facebooks-role-genocide-myanmar-new-reporting-complicates-narrative.

When coupled with keyword squatting and sockpuppet accounts, amplifying wedge issues to exploit cultural vulnerabilities disproportionately harms historically marginalized groups. Across each of these campaigns, we described how white supremacists, "trolls," and partisan pundits used the materials crafted by manipulators to sow dissent and, sometimes, target groups that are unable to refute accusations against them. Most significantly when keyword squatting, manipulators steal the voice and tools of representation afforded to minoritized groups through social media. This particular tactic will be the most difficult to detect and debunk. *We advise journalists to seek out an abundance of corroborating evidence when reporting on the actions of social media accounts, and whenever possible, verify the identity of account holders. We suggest that newsrooms invest more resources in information security, including creating a position or desk to vet chains of evidence through analysis and verification of metadata for evidence of data craft.*

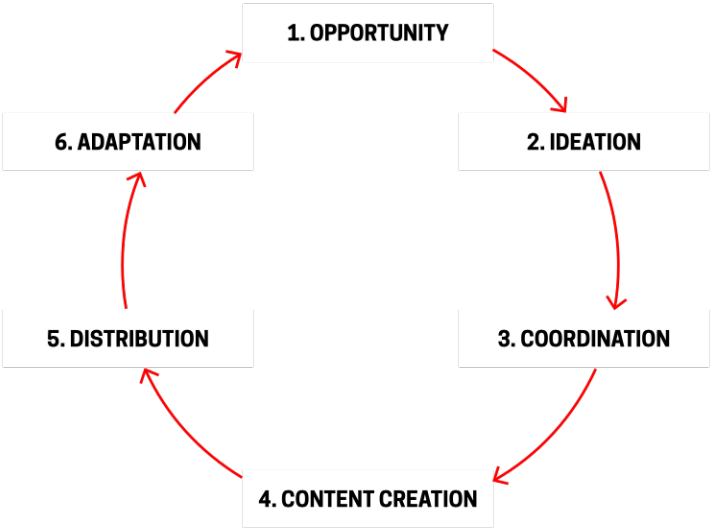
Learning the tactics of source hacking is a starting point for understanding manipulation campaigns and for designing platforms that can defend against them. *Platform companies must label manipulation campaigns when they are identified and provide easier access to metadata associated with accounts.* Facebook has already begun doing so by issuing blog posts about "coordinated inauthentic behavior." As we have started here, more can be done to create a taxonomy of types of campaigns. Terms like viral sloganeering and keyword squatting should be used instead of inaccurate terms like "trending" or "trolling," which are not specific enough. Further, evidence collages and forged leaks should be treated with the utmost suspicion, where swarms are likely pushing a specific narrative.

Social media has matured quickly over the last decade, and united together, social movements have produced impressive and lasting social change. However, society

is only as stable as the communication technology at its foundation. Today, there are many cracks in the wall and efforts to prevent the damage caused by disinformation will only succeed if there are stronger supports put in place, culturally, technologically, and legally. We have shown throughout this report that media manipulators are fighting an information war, enhanced by technology, and much of it is completely legal. It is remarkable how these technological innovations, which have driven the cost of global communication to a minimum, have also produced so much concern about the future of democracy. The way people remake technology, both good and bad, is consequential for how future technologies will be shaped.[80] Technologies do not change on their own nor do they alone support social change. Innovation is not inevitable. Instead, humans are at the center of socio-technical systems and, as such, human dignity must form the center of decisions-in-design.

80 Donald MacKenzie and Judy Wajcman, eds., *The Social Shaping of Technology*, 2 edition (Buckingham Eng. ; Philadelphia: McGraw Hill Education / Open University, 1999).

Appendix 1: Source Hacking Threat Model



1. OPPORTUNITY
When a crisis occurs and the public begins looking for information using specific keywords before official news verification, the opportunity for manipulation presents itself.

2. IDEATION
When an active news event gets the attention of crowdsourced, loosely coordinated groups begin to investigate and seek evidence. The goal of media manipulators is to influence these investigations and public conversations, customizing their implanted ideas in response to vulnerabilities and opportunities to seed disinformation.

3. COORDINATION
Media manipulators start a discussion in private or anonymous online spaces. These conversations routinely occur on messenger apps, image boards, private forums, or Internet Relay Chat (IRC). Off-site and cross-platform coordination neutralizes individual platform ability to effectively identify and contain the spread of emerging disinformation. During the coordination phase, media manipulators may also identify journalists or news organizations who they believe may be susceptible to intervention.

4. CONTENT CREATION
The media manipulators quickly create content consisting of text posts, collages, memes, or videos. The content is packaged for sharing as a blog, image, or video. The most effective content is usually purposeful disinformation combined with some form of objectively verifiable evidence. At this stage, the creators and origin point of the disinformation content are intentionally masked or filtered out in an effort to make it appear organic.

5. DISTRIBUTION
Coordinated action to leverage known knowledge gaps and algorithmic vulnerabilities can occur in several ways:

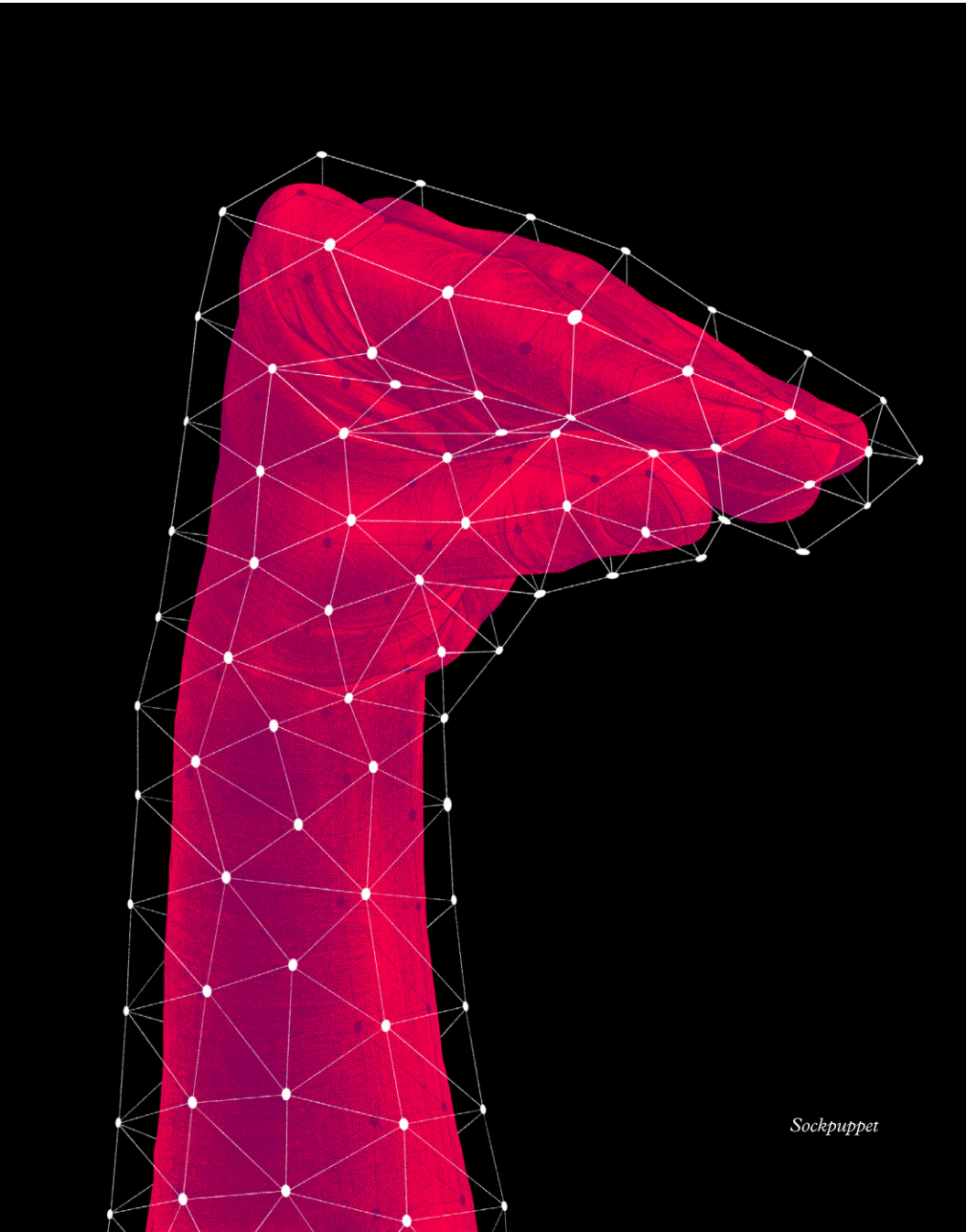
a. Create content for search terms that are highly relevant to the event, but until that moment had low algorithmic signals. During a breaking news event, manipulators will use unique keywords involving place, name, or businesses, etc., to infect a trending topic with disinformation.

b. Content can be stylized if the platform prefers specific types of posts. This could mean hyperlinking to disinformation with an attractive thumbnail image (Facebook and Twitter), placing packaged disinformation on a blog (Medium and WordPress) or video (YouTube), or combining evidence into an evidence collage (Instagram and Twitter).

c. Strategic coordinated distribution involves sharing content widely, but avoiding spam filters. Media manipulators will rely on coordination in back channels so that geolocation and posts are distributed across place and time using a mix of seasoned (abandoned accounts available for repurchase), automated (bots or purchased retweets, views, or likes), and sockpuppet accounts.

d. To move disinformation into the mainstream, it is necessary to get influencers' and/or journalists' attention. Media manipulators can ask for shares, retweets, likes, and comments publicly or in private messages. Public replies and comments are used to raise provocative questions, usually with loaded keywords, which serve a dual purpose of reaching the influencers/journalists and their followers.

6. ADAPTATION
Active manipulation campaigns are dynamic and responsive to new developments and public interest across specific topics or keywords. These campaigns iterate as amateur and official investigations progress, keywords of interest change, and new vulnerabilities are discovered.


Sockpuppet

Acknowledgments

Previous iterations of this work were enhanced by the research and guidance of Data & Society's Media Manipulation team, including danah boyd, Becca Lewis, Kinjal Dave, P. M. Krafft, Matt Goerzen, and Patrick Davison.

Data & Society

Data & Society is an independent nonprofit research institute that advances new frames for understanding the implications of data-centric and automated technology. We conduct research and build the field of actors to ensure that knowledge guides debate, decision-making, and technical choices.

www.datasociety.net
@datasociety

Illustrations:
Sockpuppet by Jim Cooke

DATA&
SOCIETY

www.datasociety.net
@datasociety