Political Communication

Routledge
Taylor & Francis Group

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/upcp20

Disinformation by Design: The Use of Evidence Collages and Platform Filtering in a Media Manipulation Campaign

P. M. Krafft & Joan Donovan

To cite this article: P. M. Krafft & Joan Donovan (2020) Disinformation by Design: The Use of Evidence Collages and Platform Filtering in a Media Manipulation Campaign, Political Communication, 37:2, 194-214, DOI: 10.1080/10584609.2019.1686094

To link to this article: https://doi.org/10.1080/10584609.2019.1686094

Published online: 05 Mar 2020.

Political Communication, 37:194–214, 2020
© 2020 The Author(s). Published with license by Taylor & Francis Group, LLC
ISSN: 1058-4609 print / 1091-7675 online
DOI: https://doi.org/10.1080/10584609.2019.1686094

Routledge
Taylor & Francis Group

Disinformation by Design: The Use of Evidence Collages and Platform Filtering in a Media Manipulation Campaign

P. M. KRAFFT and JOAN DONOVAN

Disinformation campaigns such as those perpetrated by far-right groups in the United States seek to erode democratic social institutions. Looking to understand these phenomena, previous models of disinformation have emphasized identity-confirmation and misleading presentation of facts to explain why such disinformation is shared. A risk of these accounts, which conjure images of echo chambers and filter bubbles, is portraying people who accept disinformation as relatively passive recipients or conduits. Here we conduct a case study of tactics of disinformation to show how platform design and decentralized communication contribute to advancing the spread of disinformation even when that disinformation is continuously and actively challenged where it appears. Contrary to a view of disinformation flowing within homogeneous echo chambers, in our case study we observe substantial skepticism against disinformation narratives as they form. To examine how disinformation spreads amidst skepticism in this case, we employ a document-driven multi-site trace ethnography to analyze a contested rumor that crossed anonymous message boards, the conservative media ecosystem, and other platforms. We identify two important factors that filtered out skepticism and contested explanations, which facilitated the transformation of this rumor into a disinformation campaign: (1) the aggregation of information into evidence collages—image files that aggregate positive evidence—and (2) platform filtering—the decontextualization of information as these claims crossed platforms. Our findings provide an elucidation of "trading up the chain" dynamics explored by previous researchers and a counterpoint to the relatively mechanistic accounts of passive disinformation propagation that dominate the quantitative literature. We conclude with a discussion of how these factors relate to the communication power available to disparate groups at different times, as well as practical implications for inferring intent from social media traces and practical implications for the design of social media platforms.

Keywords disinformation, Alt-right, 4chan, tactics, media manipulation

P. M. Krafft is a Senior Research Fellow at the Oxford Internet Institute. Dr. Krafft studies beliefs, ideology, and institutions in the information society, and works and organizes in the area of public interest computing. Joan Donovan is the Director of the Technology and Social Change Research Project at Harvard Kennedy School's Shorenstein Center on Media Politics and Public Policy. Dr. Donovan researches media manipulation, disinformation, and adversarial media movements that target journalists.

Address correspondence to P. Krafft, Oxford Internet Institute, University of Oxford, 1 St Giles', Oxford OX1 3JS, UK. E-mail: p.krafft@oii.ox.ac.uk

Data collection and analysis was conducted at Data & Society Research Institute. Writing was conducted at Data & Society, the University of Washington, and Harvard University.

Introduction

Rumors can in principle become established facts. At least in the case when substantial and irrefutable corroborating evidence is marshaled to support them. Yet the process of gathering and distributing evidence is difficult and fragile. We examine how this process of information gathering and distribution can be corrupted by media manipulators, i.e. those seeking to seed doubt or disinformation into public conversations for various reasons. Synthesizing the definitions of Jack (2017) and Starbird (2019), we define the activity of disinformation as creating and distributing intentionally deceptive content. Rumors often spread in ambiguous situations (Allport & Postman, 1946). For this reason, people who spread disinformation can leverage the ambiguity of a situation to their advantage. But why would disinformation be believed, rather than viewed with suspicion or taken as uncertain, when a situation is ambiguous? In the present work we analyze disinformation tactics used by various media manipulators to demonstrate how they co-opt media artifacts and platforms to diminish skepticism about their claims.

We examine this question in the context of disinformation arising from 4chan. Launched in 2003, 4chan hosts both messaging and image board forums that serve to generate and categorize user conversations. One notorious board, 4chan/pol/, has been a well-known organizing hub for white supremacists and the U.S. far-right.[1] Several previous studies have examined how information propagates from 4chan to other areas of the contemporary hybrid media ecosystem, e.g. to Twitter or major news outlets, through mechanisms such as “trading up the chain” (Marwick & Lewis, 2017). Trading up the chain involves people formulating and popularizing narratives in contexts like 4chan message forums, and then deliberately trying to increase the visibility of those same narratives through more mainstream media actors—from far-right blogs and forums to conservative media personalities on Twitter or YouTube to television media outlets like Fox News.

One partial explanation for why disinformation forms and spreads across conservative media ecosystems is that these spaces are relatively homogeneous “echo chambers,” cut off from diverse political viewpoints that challenge information asymmetries and disinformation (Benkler, Faris, & Roberts, 2018; Marwick & Lewis, 2017). According to some such views, people consuming content from these venues spread identity-confirming disinformation narratives as a component of meaning-making through their engagement, even after being exposed to conflicting narratives (Marwick, 2018). In what follows, we investigate the extent to which challenges to disinformation exist in the far-right digital media ecosystem and what additional factors should be considered to understand how disinformation propagates in these contexts. To do so, we revisit a disinformation case that occurred during the 2017 Charlottesville Unite the Right Rally (henceforth, UTR). On August 12, 2017, a crowd of counter-protesters amassed to oppose the hateful messages of a loosely organized far-right movement unified by white supremacist and misogynist ideologies (Hawley, 2017; Wendling, 2018). In a tragic moment of violence that would become the climax of the chaotic weekend, a white supremacist drove his car into the crowd of counter-protesters, killing activist Heather Heyer and injuring dozens of others. In the immediate wake of the attack and prior to the public release of the attacker’s identity, a disinformation campaign ensued in which “Joel V.,” a left-leaning musician and activist living in Michigan and not present at the event was falsely accused of perpetrating the attack.

We document the context surrounding the formation and spread of this UTR disinformation campaign. Contrary to the homogeneous space that 4chan is sometimes described as, we observed widespread heterogeneity of beliefs and contestation of the claims that became the dominant disinformation narrative. Given the presence of skepticism and dissent among forum participants, appealing to identity-confirmation alone as a mechanism falls short of a complete explanation of how this disinformation campaign spread. The situation also diverges from the simple stories told by typical quantitative models of attitude and belief dynamics, in which persuasion occurs through passive exposure (cf., Castellano, Fortunato, & Loreto, 2009; Friedkin & Johnsen, 2011). Rather than passive exposure to identity-confirming narratives driving disinformation, we find that an arsenal of tactics that exploit the structure of social media is deployed to spread disinformation in the case we study.

We document two additional factors that contributed to the disinformation in this case. While there is evidence to support that identity-confirmation is a component of why certain participants on 4chan spread disinformation, other major factors also arise. First, we broadly sketch out how media artifacts called evidence collages are strategically constructed to reinforce and garner support for disinformation campaigns by eliminating competing claims. Second, in the process of the evidence collage we study being reposted on other platforms and elsewhere on the web, the context of its origins and prior counter-narratives were filtered out. We characterize the latter effect as platform filtering, which is a source hacking (Donovan & Friedberg, 2019) technique related to a form of information flow curation (Thorson & Wells, 2015) through a decentralized gatekeeping-like tactic (cf., Shoemaker & Vos, 2009). We argue that these factors are especially important to consider because they expose the clear intentionality of the disinformation campaign. We find that disinformation did not just spread on its own because it affirmed people's identities, rather it was the result of an intentional strategy to move the disinformation campaign through the larger media ecosystem.

We conclude by connecting our findings to several contemporary topics in political communication. On the practical side, we discuss how our analysis provides evidence for intent in disinformation formation and distribution, which could have broader implications regarding identifying intent from social media activity, and we offer suggestions for furthering transparency and accountability mechanisms. On the theoretical side, we explore how disinformation tactics inform understandings of communication power (Castells, 2013), including how disinformation relates to the fragmentation of conversational contexts across platforms. Theoretically locating disinformation campaigns as occurring within sociotechnical systems, which integrate both social factors and technical features, is crucial for understanding and mitigating disinformation's impact on society. Both people and platforms play roles in disinformation spreading.

Background

Disinformation Tactics

To study disinformation tactics, Bennett and Livingston (2018, p. 124) call for cross-platform analysis, stating: "While the origins of much, and perhaps most, disinformation are obscure, it often passes through the gates of the legacy media, resulting in an 'amplifier effect' for stories that would be dismissed as absurd in earlier eras of more effective press gatekeeping." A handful of cross-platform empirical analyses explore these obscure origins, and disaggregate the paths of amplification by mainstream media.

Other researchers studying disinformation tactics focus on automated techniques of computational propaganda, defined as "algorithms, automation, and human curation to purposefully distribute misleading information over social media networks" (Barberá et al., 2018; Woolley & Howard, 2016, 2018). In these cases, effective disinformation campaigns infiltrate multiple web spaces and platforms and then use a range of customized attacks to spread disinformation across networks. Our research is less concerned with identifying complex disinformation campaigns involving computational propaganda, but instead focuses on "low tech" or "no tech" exploits within open information environments.

Open web forums are often used as basecamps for coordinating and planning disinformation campaigns. Both Zannettou et al. (2017) and Marwick and Lewis (2017) identify 4chan's image boards as a source of disinformation that propagates to other platforms. Marwick and Lewis state:

> Pizzagate illustrates a particularly wide range of related phenomena: a conspiracy theory developed and grew within online networked communities; misinformation and 'fake news' spread virally through social media; believers undertook collaborative efforts that ultimately reinforced their previous views; an individual was quickly moved to action through exposure to false information; and coverage from the mainstream media simultaneously amplified the theory and left its believers feeling disenfranchised.

Although Benkler et al. (2018) argue that online disinformation is comparatively less impactful on societal outcomes than major media organizations, they also confirm the general pattern identified by Marwick and Lewis (2017), emphasizing how trading up the chain for pieces of disinformation—from sites like 4chan—is crucial to having further reach across networks.

Much research on how disinformation moves across platforms focuses on text-based narratives. Our research pivots toward images as a medium of visual political disinformation (Griffin, 2015). There is substantial research on visual memes and online communities (Phillips, 2015, 2018; Wall & Mitew, 2018). Other scholars have emphasized the importance of visual propaganda as a tool for enabling online extremism (Dauber & Winkler, 2014), and more specifically, for the radicalization of white-ethnoterrorists (Waltman, 2014). We use Mathison's (2009) four aspects of credibility of visual misinformation to analyze the evidence collages spread in the wake of UTR. Mathison (2009) defines credibility as "typically established through a combination of coherent presentation, a quality of authenticity of legitimacy, relevance to the viewer, and a sense that the information being represented is in its original context."

In contemporary accounts of disinformation campaigns, little attention is paid to the multifaceted roles of multiple exposures to disinformation, the challenge of recruiting media outlets with wider viewership, how these factors mesh with the passive identity-confirming characteristics of media consumers, and what mechanisms help overcome impediments to disinformation spreading. Our analysis illustrates that in times of high uncertainty, visuals can recalibrate a sense of proof by mimicking the appearance of a credible source. When coupled with identity-confirming content and the openness of platforms, claims to truth can be channeled away from challenges and expressions of skepticism.

Rumors and Uncertainty

A rumor can be defined as "an unverified proposition for belief that bears topical relevance for persons actively involved in its dissemination" (Rosnow & Kimmel, 2000). The classical literature on rumors points to two preconditions for rumoring:

ambiguity and importance (Allport & Postman, 1946). These researchers argue that for a rumor to occur the situation must be uncertain, and people must care about the content of the rumor. Because uncertainty or lack of information coupled with personal connection is a defining precondition for a rumor to spread widely, rumors are frequently observed during chaotic scenarios, such as disasters or crises when little authoritative information is available (Bordia & DiFonzo, 2004; DiFonzo & Bordia, 2007; Simon, Goldberg, Leykin, & Adini, 2016; Spiro et al., 2012). Crisis creates a context whereby multiple conflicting narratives can simultaneously co-exist while directing public attention and galvanizing political support.

Another classical theory of rumoring behavior ties rumors to meaning-making by supposing that participating in rumoring reflects participants efforts at collective sensemaking (Shibutani, 1966). Central to the collective sensemaking theory is the role that doubt and skepticism play in spreading rumors. Researchers have documented the presence of collective sensemaking during disaster and crisis contexts, and have expanded on the role of social media platforms in amassing content and increasing user activity (Arif et al., 2017; Dailey & Starbird, 2015; Heverin & Zach, 2012; Maddock et al., 2015).

During collective sensemaking, people may create theories, introduce evidence, and process evidence. Narratives are supported and debunked, and gain and lose popularity over time. In studying whether crowds can self-correct false rumors, researchers have noted substantial shifts of opinions held by individuals (Arif et al., 2017), and greater expression of uncertainty has been documented when individuals are confronted by false rumors (Mendoza, Poblete, & Castillo, 2010). While some work suggests that misinformation can be more persistent than corrections (Shin, Jian, Driscoll, & Bar, 2017; Starbird, Maddock, Orand, Achterman, & Mason, 2014), patterns differ between rumors (Arif et al., 2017). While debunking rumors appears to influence group dynamics, the sheer volume of a rumor can shift whether it is corrected (Friggeri, Adamic, Eckles, & Cheng, 2014). Recent work has argued that multiple actors and distinct communities participating in collective sensemaking in parallel may also partly explain in some cases why false rumors are so persistent (Krafft, Zhou, Edwards, Starbird, & Spiro, 2017).

This literature on rumors suggests focusing on the ancillary participants in conversations where disinformation appears and analyzing the conversational contexts surrounding disinformation. Scholars studying rumors often expect to find substantial skeptical engagement during highly uncertain and contested situations. We should therefore expect that rather than disinformation flows involving many people who either believe disinformation content or ignore it, participants who question or challenge disinformation flows are important to consider in studying disinformation campaigns. How do actors who spread disinformation overcome skeptical remarks and other challenges? To answer this question, we must address how power and influence operate to shape the sociotechnical systems where rumors and disinformation campaigns circulate.

Decentralized Communication Power

Our case study bears on theories of communication power. More precisely, we focus on networked social movements, which includes the capacity for individuals and groups to network and coordinate to advance a particular outcome (Castells, 2013). As the hybridity of the media ecosystem has increased, the nature of communication power of both elites and non-elites has shifted (Chadwick, 2017). Computational infrastructure spurs personalized campaigns that isolate audiences to gain individually targeted influence (Nadler, Crain, & Donovan, 2018; Tufekci, 2014). Algorithmic filtering may create filter bubbles that isolates people from ideological challenges (Bozdag, 2013). State actors have devoted substantial resources to creating and promoting propaganda online (Woolley & Guilbeault, 2017; Woolley & Howard, 2017). Owners of digital platforms derive political power from control over content moderation and curation (Wallace, 2018).

Theories of non-elite power in new media environments exist but have often focused on modeling the achievement of positions of power through rapid crowdsourced growth. These types of theories of new media environments resemble theories of centralized power with new routes to becoming elite. Meraz and Papacharissi (2013) conducted influential empirical work on network gatekeeping in the social media context, but focused on the power associated with central network roles. Sayre, Bode, Shah, Wilcox, and Shah (2010) and Meraz (2009) examined the agenda-setting abilities of popular users of new media platforms. For example, Sayre et al. (2010) emphasize: "What has made YouTube a new force to be reckoned with is summed up in its marketing slogan: Broadcast yourself." Leavitt and Robinson (2017) examine collaborative filtering during crisis events, focusing on the tensions at play when the crowd curates information. While valuable and nuanced in other ways, these works, in-so-far as the need for new theories of media effects is concerned, essentially replace the elite actor with "the crowd." The crowd empowers users by its interaction with platform design that promotes popular content. Offering a more flexible model, Wallace (2018) proposes a framework for analyzing digital gatekeeping, wherein access to information, selection criteria for information, and a choice of publication space define a process of gatekeeping available to any actor. Thorson and Wells (2015) offer that the gatekeeping metaphor should be replaced by a metaphor of "curated flows," wherein each user has access to varied flows of information that are curated in different ways—whether by journalists, strategic actors, algorithms, or personal choices. Both frameworks are helpful but still fall short of capturing unique aspects of decentralized disinformation campaigns, including the degree to which these campaigns are coordinated and the degree to which they infiltrate publics rather than spring from or replace publics.

In our case study, we analyze the activities of people with potentially little inherent positional advantage deriving from their network connections or popularity on a platform. Metaphors and theoretical models for political communication have yet to catch up to these tactics of decentralized communication power. From what source do disorganized or loosely organized actors who have little inherent capacity for influence from their network positions derive power? Castells (2012) and Donovan (2018) argue that the infrastructure of the Internet allows for decentralized command and control of information flows, and that the technical design of platforms is reflected in the communication structure of networked social movements. When networked social movements "play the algorithm," it is primarily to spread information or to influence the behavior of an algorithm by coordinating groups of people to share specific media artifacts (Monterde & Postill, 2014, p. 433). Eventually certain artifacts independently move across networks and platforms, effectively splintering from the initial push by movement organizers or particular individuals who rely on their status as online influencers to initiate the spread of a campaign (Tufekci, 2014).

In terms of communication power, platform companies centralize power, while the web decentralizes it. Corporations can decide if content or users are banned or promoted, but the crowd also impacts the creation and flow of content. Our case study documents how an authorless piece of content can become an authoritative source when it leverages the design of platforms to filter out dissent and skepticism, while also trading up the chain to gain legitimacy.

Methodology

Because we are interested in establishing a detailed account of the mechanisms of media manipulation tactics, we rely on a case study-based methodology (Ragin & Becker, 1992). To set the boundaries and criteria for choosing a case study, we consider the case of a single disinformation narrative that exemplifies the dynamics of trading up the chain. We choose a case from a crisis situations because crisis situations are marked by uncertainty and rumors, and our research question concerns disinformation tactics in the presence of uncertainty and skepticism, particularly when authoritative information is scarce.

We use a document-driven multi-site trace ethnography (Geiger & Ribes, 2011) to analyze the case, and rely on a "follow the actors" approach to understanding distributed sociotechnical systems and digital traces of behavior. By doing so, our research reconstructs the activities and relationships of disparate actors. Our approach involves systematically collecting web data related to our case, reconstructing a timeline of events relevant to the spread of the case of disinformation we study, and analyzing the conversational contexts where the disinformation campaign travels.

Case Selection

On August 12, 2017 during the UTR rally in Charlottesville, Virginia, a crowd of counter-protesters amassed to oppose the hateful messages of the event—constituted by a loosely organized far-right movement unified by white supremacist and misogynist ideologies (Hawley, 2017; Wendling, 2018). Our case traces a false claim that the perpetrator of the car attack at UTR was a counter-protester of the event identified as Joel V. A notable aspect of this disinformation was that it spread in part through a set of images that mixed together pieces of evidence supporting the disinformation campaign targeting Joel V. Our analysis focuses on the creation and distribution of one image. By focusing on one image variant, we are able to bound the data collection procedure and better identify and track the formation of disinformation.

Our research focuses on the twelve hours after the time of the car attack. During this time, politically-motivated rumors percolated about the identity of the car driver. These rumors were fueled by 4chan's and similar platforms' crowdsourced investigation, reminiscent of misguided online efforts during the Boston Marathon Bombing manhunt (Starbird et al., 2014). At the peak of this rumoring, conservative alternative media websites posted stories falsely identifying Joel V. as the perpetrator. Meanwhile, Joel and his family faced harassment via social media, e-mail, and telephone until they eventually had to leave their home.

Tracing Methods

We collected messages and conversations containing discussion of the car attack rumor from 4chan, 8chan, Reddit, Twitter, Discord, web blogs, and web forums.[2]

We chose this selection of platforms because of their established importance by previous researchers of the far-right, and for the role they play in trading up the chain dynamics. We leveraged initial 4chan data collected at the time of the Charlottesville rally, and systematically collected data from 4chan and the other websites from September to December 2017. We analyzed this data by creating a tabular representation of key claims made in each source, comparing time stamps of each event, and referencing archival sources such as the Way Back Machine and Archive.is to verify the data. The uniqueness of the full name of the false suspect, Joel V., allowed us to easily identify posts relating to this rumor.[3] We collected retrospective 4chan and 8chan data by searching through the thread indexes of those websites, which were archived on http://archive.is/ and http://archive.4plebs.org/, for threads that were related to UTR. We downloaded Reddit data from http://files.pushshift.io/reddit/, and searched through the downloaded data from the date of the attack for instances of the name Joel V. We collected Twitter data through the website's search function, searching for instances of the name Joel V. on the date of the attack. Data from Discord was available from https://discordleaks.unicornriot.ninja. We also conducted web and news searches via Google using the full name of Joel V. as a search term to identify blogs and forums mentioning the rumor on the date of the attack, and we conducted Google reverse image searches on the evidence collage we study. We supplemented these data sources with links identified in Politifact and Snopes articles related to the attack. We then conducted a systematic reconstruction of a timeline of one variant of the rumor by tracking the timestamps of messages and first occurrences the rumor variant.

The online platform 4chan has played a major role in propelling disinformation and leveraging other online influence campaigns (Hine et al., 2017). 4chan's /pol/ board is part of a broader network of forums associated with white supremacist and far-right organizing (Marwick & Lewis, 2017), including 8chan/pol/ and far-right subreddits. A substantial portion of the content on these sites is explicitly racist, anti-Semitic, and misogynistic. Embedded within this context is a range of activities oriented toward shaping compelling public narratives. One notable activity in these forums is called digital sleuthing, or crowdsourced investigations. Forum members use existing public evidence and search engines to forge allegations and pinpoint suspects before an official account is given by the police. In the case of "Joel V.," we document how activities that began with crowdsourced investigations and rumoring transformed into an explicit disinformation campaign.

A distinctive feature of communication in 4chan and similar forums is the use of visual media to summarize ideas and narratives. Participants create images of information and aggregate these images into evidence collages, which can also be used to spread disinformation, when used to slander a group or individual. The rumor variant we study takes the form of one of these digital artifacts, an evidence collage that aggregated the information that Joel V. perpetrated the Charlottesville car attack. Other variants collaging the same pieces of evidence existed in parallel with this collage, which illustrates how common this communication tactic is and illustrates the representativeness of the collage we study. To maintain the tractability of the case study, our investigation is restricted to one collage that appeared frequently.

Another distinctive feature of 4chan is that it places a limit on the length of threads, and threads disappear after they become too long. This feature creates a dynamic environment and provides an opportunity for a new "original poster" to initiate the discussion around the subject in a new thread, which opens the possibility of narrative reframing.[4] These thread dynamics, including the rapid filtering of information, became important in understanding how the disinformation campaign was formed.

Three other aspects of evidence collages and the Joel V. rumor led to our subject being uniquely suitable for rigorous analysis. The evidence collage was built iteratively over time, which allowed us to trace the visual representation of the evolution of the Joel V. rumor. Components of the image had unique identifiers that allowed us to verify whether or not the components were arranged iteratively by a single creator. Whoever took screenshots of two key pieces of evidence from the web that were integrated into the collage at different times did not crop out the battery level or notifications on their phone. The spacing between images and red guiding lines in the collage also have unique locations. These digital forensics allow us to identify plausible first occurrences of the Joel V. rumor and construct an approximate timeline from the evidence collage components.

Rumor Timeline

In this section we discuss the trajectory we observe of the Joel V. rumor—how this rumor formed and spread—and the evidence collage we use as a focus of our study. A summary of the reconstructed timeline appears in Table 1. Our timeline focuses on the early crowd investigation behavior on 4chan to illustrate how the evidence collage was assembled. The effort began when a high-resolution photo of the back of the vehicle used in the attack, which included the vehicle's license plate, appeared on 4chan in an untitled thread at 3:07pm on August 12. Soon after the photo appeared, a user posted that they "Ran the plates … " with an image of results from a website called Search Quarry showing the registered driver as Jerome V., Joel's father. This image would eventually appear in the

Table 1
The Joel V. rumor timeline

Time	Event
1:45pm	Unidentified individual attacks crowd with vehicle
3:07pm	High-resolution vehicle license plate photograph appears on 4chan
3:17pm	First post of Search Quarry image on 4chan
3:36pm	Partial evidence collage appears on 4chan
3:55pm	Joel V. "confirmed" as suspect by 4chan
4:25pm	Partial collage and separate Facebook pictures posted to 4chan, and Joel-as-leftist narrative solidifies in new thread: "CHARLOTTESVILLE RAM WAS A COMMIE"
4:30pm	Partial collage cross-posted to Reddit's /r/conspiracy and /r/TheDonald
4:36pm	Full collage appears on "CHARLOTTESVILLE RAM WAS A COMMIE" 4chan thread
(ongoing)	Full collage and other instances of Joel V. rumor appear on Twitter and other websites
5:13pm	Joel V. posts on Facebook saying he's not the perpetrator of the attack
6:10pm	James Fields identified by 8chan as more plausible suspect
8:00pm	Washington Post reports police officially identify James Fields as suspect

final form of the collage we study. The forum participants would also eventually discover evidence via Facebook that the car was given to Joel and that Joel had left-leaning politics. Similar crowdsourced investigations were happening in parallel on other forums, and other collages were created elsewhere.

Breaking down the investigation within its first hour illustrates how quickly the rumor evolved and spread. At 3:36, a segment of the collage was posted in the same untitled thread with discussion still focusing on Jerome. At 3:39, a person identified Joel via a LinkedIn page as possibly Jerome's child, leading to further speculations surrounding Joel. At 3:58, a user found a picture on Facebook indicating that Joel had received the car used in the attack as a birthday gift several years ago, with the exclamation: "HOLY FUCK GUYS WE FOUND HIM." Nearly simultaneously, at 3:55 the same image was posted (via a different screenshot, so likely by a different person) in a new thread on 4chan, entitled "SUSPECT CONFIRMED." Concurrently, there was a thread called "Drivers name is Jerome" started at 3:19 with a progression of evidence pointing to Joel.

In the thread "SUSPECT CONFIRMED" the forum participants began to explore Joel's public photos online, which revealed Joel to be outspoken andleft-leaning. The forum participants became excited about the idea that a leftist might have run into the crowd of counter-protesters, potentially as a "false flag" attack. Visual aggregations of evidence in the form of collages began to appear in this thread. The narrative of Joel as a leftist perpetrator coalesced further after the creation of a new thread, "CHARLOTTESVILLE RAM WAS A COMMIE" at 4:25pm. This thread began with a segment of the full evidence collage, which itself appeared in full soon after, at 4:36pm in this same thread. Once these visual artifacts were created, the narrative became easily communicable and easily shareable. We observed both later posts of the theory that Joel was the driver and the image itself in a variety of outlets, which is indicative of trading up the chain. Conservative media sites, such as GotNews and Puppet String News, shared the story soon after it took shape on 4chan, and it also appeared on other platforms such as FreeRepublic, Zerohedge, Reddit, Twitter.

The platforms most impacted by the Joel V. disinformation campaign were alternative conservative media websites. For instance, the GotNews article "Breaking: Charlottesville Car Terrorist is Anti-Trump Open Borders Druggie" unequivocally adopted and rapidly shared the narrative spun on 4chan. In a ZeroHedge article that was tracking the developments of the attack, one user posted in a comment: "This outfit here is saying it's … Joel V[redacted] and they're calling him an anti-trump guy FWIW," and provided a link to the GotNews article. Another user replied saying: "Once the press finds out this guy is a lefty this story will soon vanish," a post which received 139 upvotes. The broader ZeroHedge article has since garnered 773,346 views and 1,631 comments at the time of this writing. The GotNews article received enough attention that its owner became one of the defendants named in a defamation lawsuit filed by Joel V. When the collage appeared on Twitter, these tweets were quickly challenged by other users on the platform. Nevertheless, the posts on Twitter helped spread the rumor as users tweeted in the replies to breaking news from mainstream media outlets.

Analysis

Our research question pertains to how disinformation spreads despite expressions of dissent or uncertainty. We first document how our case fits next to existing explanations of disinformation flows. We then confirm that skeptical and hostile reactions appear in conversations where disinformation is present. We finish our analysis by detailing how two key factors—evidence collages and platform filtering—help explain the persistence of disinformation sharing despite the presence of skepticism in the case we study.

Identity-Confirming Content

Benkler et al. (2018) argue that the right-wing media ecosystem is especially susceptible to disinformation because there are weaker norms of truth-seeking, fewer network connections to center media, and a greater propensity for identity-confirming content. Similar arguments have been made concerning ideological homogeneity across conservative media sites and discussion boards (Marwick & Lewis, 2017). In our case study, we find direct evidence for a propensity for identity-confirming content. Some participants began their investigation with the explicit goal of identifying the driver as a black person or as a counter-protester. For example, a participant in one thread called "Drivers name is Jerome" stated: "Dude that last name is African." As another example, participants in the forum later became fixated on Joel V.'s left-leaning views, which they wanted to try to use to defame counter-protesters from the left and discredit the mainstream media's account of the events. Although aspects of an identity-confirmation account during trading up the chain are validated by our case study, the story is also more complicated than a simple picture of like-minded participants on 4chan and other forums. Another poster in the thread titled "Race of the driver," challenged the above example identity-confirming statement, writing, "Alright guys listen up, it doesn't matter if the driver was white or black because he could possibly be a right winger or an Antifa member regardless of race so stop fucking arguing over the drivers race because he's a domestic terrorist either way."

Skepticism and Uncertainty

Consistent with a sizable literature documenting collective sensemaking activities and their associated expressions of uncertainty surrounding rumors, we observe a variety of skeptical posts, even as it was created on an explicitly far-right 4chan forum. Unlike some cases of disinformation, the Joel V. narrative does not appear to have been created by a single person. The process of formalizing the narrative involved multiple accounts sharing pieces of evidence from various sources with some degree of legitimacy. Probably in anticipation of skepticism and challenges, participants in the formation of the narrative appear to have seen a need to appeal to legitimate sources of evidence, such as bureaucratic records of the vehicle registration and Joel's own Facebook page. The license plate number, the vehicle registration, the identification of Joel's Facebook account, and the discovery of the car used in the attack in Joel's photos all served as important visual corroborating evidence. These components played a role both in the initial discussions about the disinformation, and later in the evidence collages that were created to spread it. Still, dissenting participants on 4chan questioned these pieces of evidence individually, asserting for example that the car had been stolen or that the registration was illegitimate because it was issued by the state of Michigan while the plates were from Ohio. In all the 4chan threads we analyzed whenever Joel was discussed there were also dissenting voices. Participants also challenged the narrative in the other venues where it later spread. Given these findings, the remainder of our analysis will shift toward our main research question: How did the Joel V. continue to circulate alongside persistent skepticism from within 4chan and other venues?

Evidence Collages

Images are media artifacts that can easily be shared across threads or platforms. Online culture on social media and message boards provides a context where sharing images is normalized. We argue that the use of visual evidence collages represents a key strategic element in the formation and spread of disinformation.[5] The familiarity within 4chan of evidence collages provided an opportunity for a single actor to aggregate information into a compelling and easily digestible and shareable format. Forum participants were able to craft the image according to the story they wanted to share, while also opening up major political opportunities for further media manipulation. As Mathison (2009) stipulates, visual images are compelling when they are coherent, legitimate, relevant, and original. Although multiple actors participated in the overall formation of the Joel V. narrative, our analysis suggests that a single actor engaged in iterative attempts to shape the image in the collage into a compelling piece of visual media accusing Joel V. A partial version of the collage was included in the initial post of the "CHARLOTTESVILLE RAM WAS A COMMIE" 4chan thread. The same actor, identifiable by an anonymized user ID string in the thread, immediately posted several other images and explicitly exclaimed: "SPREAD THIS EVERYWHERE. DON'T LET THE MEDIA BLAME THIS ON US." Minutes later, the same actor posted the first instance we observe of the collage, which built directly upon the earlier partial collage in that user's initial post, as is identifiable by the digital forensics in the image. Because the evidence collage assembles interrelated screenshots that can each be independently verified and corroborated through search engines, the media manipulator offers a coherent and seemingly legitimate false association between Joel V. and the car. To some forum participants and later viewers, the collage likely appeared relevant, authentic, and original. Also importantly to later viewers, the images contain no references to 4chan, so when they were embedded in blogs or on other platforms, their origin point was not only obscured, but also decontextualized.

Decontextualization and Platform Filtering

Bennett and Pfetsch (2018) note that the current era of political communication is marked by disrupted public spheres where social media plays a crucial role in fracturing public debate. A notable pattern in the case of disinformation we study, related to trading up the chain, is the movement of disinformation from one thread or platform to another. Decontextualization occurs when the content of a message is reproduced in a conversational context other than the one in which it first appeared, and when the prior context of that message is not reproduced in the new conversation. Decentralized communication infrastructures, like the web and platforms, foster decontextualization by allowing networks to fork conversations in different ways, creating fragmented publics. One negative effect of this design characteristic is that the opportunity for decontextualization helps spreaders of disinformation evade dissent by moving into new contexts.

To describe how decontextualization aids disinformation, we must first clarify what it means for content to move from one context to another. We define an integrated conversational context as the total information that is readily available to all the participants and observers of a communicative exchange between two or more people. Integrated conversational contexts result from public or shared records of conversations, such as those stored on digital platforms. Examples include a reply thread on Twitter, the first page of a discussion thread on a forum, or the top of a comment section on a video or article. Integrated conversational contexts are dynamic and contingent on both the user interface of a website, and the design of how information is presented across platforms. For example, when older comments on an article are pushed below a page fold, they leave the integrated conversational context of that article. Cross-platform integrated conversational contexts are possible, such as in mailing lists—all members of a mailing list receive the same messages regardless of the e-mail client they are using. Integrated conversational contexts are shaped by user interface design in three ways: (1) platform boundaries, (2) thread boundaries, and (3) page boundaries. A platform is a piece of user-facing web software that facilitates digital communication between users, such as Twitter, Facebook, or WhatsApp. There will be many integrated conversational contexts within each platform, but platforms still form important boundaries given that many platforms make their own distinct user interface choices and have distinct user bases. A thread is a group of digital messages by participants in a conversation grouped together by a user interface. A single thread, such as a forum thread on 4chan or a reply thread on Twitter, is also often broken up by page boundaries; a user must often follow a link to view the thread and may have to click several times to get the entire thread to open. Once comments are reordered by an algorithm or hidden behind links on multi-paged threads, the readily available information becomes the most easily accessible content, which receives far more attention than buried information (Hodas & Lerman, 2012; Lerman & Hogg, 2014). Crucial to the spread of information is the breakdown of bounded spaces by users that move beyond integrated contexts and can render new conversations.

Integrated conversational contexts provide an opportunity for manipulators to leverage decontextualization as information moves across boundaries. Decontextualization can remove the context in which disinformation was created, such as who created it or why it was created, and can remove the context of replies to that disinformation. Expressions of doubt or other expressed challenges to disinformation become a part of an integrated conversational context and as such become available to observers and peripheral parties to disinformation conversations. Since debunking hampers rumor sharing (Friggeri et al., 2014), we expect that visible expressions of doubt or disconfirming evidence have a similar effect. Reposting disinformation in the same thread, a different thread, or a different platform not only creates a new opportunity for exposure to that disinformation, but also simultaneously launders the content of any previous context, including any skepticism or dissent.

When a person creates a new thread on 4chan, as is required by the structure of the platform when a thread becomes too long, the dissenting voices in any previous threads are partially lost. The "CHARLOTTESVILLE RAM WAS A COMMIE" thread is an example of this phenomenon, and indeed appears to be where the "Joel-as-leftist" narrative first crystallized in 4chan. The removal of the original context of collective sensemaking provides a filtered view of the positive evidence for the rumor. When this form of redundancy occurred within 4chan, it had limited negative impact, probably in part because many of the same or similar participants from earlier threads posed the same critiques of the narrative in new threads. The platform boundary facilitated new threads replicating important pieces of the prior integrated conversational context.

An especially pernicious form of decontextualization across integrated conversational contexts occurs when a rumor crosses platform boundaries, rather than just thread boundaries. In such cases platforms behave as content and context filters, so we call this dynamic platform filtering. Consequential examples of platform filtering during the Joel V. rumor include the instances when the evidence collage was posted to platforms outside of 4chan, such as in Twitter discussions, and when the Joel V. rumor was taken up as a scoop by conservative media sites. When a reader sees only this filtered information, they lack the initial discursive context of the narrative, and may be more apt to believe it. Although behaviors such as searching to verify the information (cf., Dutton, Reisdorf, Blank, & Dubois, 2019) could mitigate this effect, other forms of legitimation that could build trust in the information also occur when trading up the chain takes place. For instance, social cues such as a retweet, a like, a share, or a repost potentially reflect a tacit endorsement of the content in question.

In the Joel V. evidence collage, only confirming evidence is presented. The narrative and evidence have been shaped into a compelling form, and there is no indication of its 4chan origins. Placing the evidence inside an image ensures that the narrative travels across platforms intact, while the surrounding text-based discussion is left behind. As a result, the primary location of narrative crystallization in this case is at the point when the evidence collage is filtered as it crosses between integrated conversational contexts, and especially platform boundaries. Platform filtering is therefore a powerful mechanism that can be exploited for strategic use by actors wishing to spread disinformation. Instead of relying on major influencers in a network to push the Joel V. campaign, media manipulators leveraged the openness of the web and the fragmentation of platforms to seed the evidence collage in new conversational contexts. Through deliberate information filtering, 4chan capitalized on the urgency of a crisis by removing skepticism while targeting audiences seeking information about the UTR event.

Theoretical Details: Context, Actors, and Roles

Our study investigates an elaboration of trading up the chain dynamics, and particularly a set of sub-mechanisms involved in an "active" version of trading up the chain (as opposed to, in the extreme, a passively conceived version that hinges purely on identity-confirmation). The active form that we identify involves sub-mechanisms of evidence collaging and platform filtering. Both of these sub-mechanisms involve particular features that set theoretical conditions for when we might expect to observe these sub-mechanisms in other cases. One key feature of our study is discussion participants actively assessing legitimacy of sources, coherence, and narrative elements. Evidence collaging can only occur when there is evidence to be collaged, and platform filtering only has the weight it carries in our case when skepticism is present.

The mechanisms of evidence collaging and platform filtering we identify involve several types of actors with different roles. The actors we identify as the strategic disinformation campaigners are those who begin the process of investigation with identity-confirming motivations, the evidence collage creator, and the individuals who move the collage across platforms. The ancillary participants who either express skepticism or participate in collective sensemaking in smaller parts are less easily associated with the key parts of the disinformation campaign in our case. It is hard to distinguish from the data we have whether the actor who create the evidence collage and the actors who strategically move it between platforms are a single person or multiple people who are actively coordinating or multiple people acting without any direct coordination in the moment. Once the evidence collage is created, this artifact lends itself to inadvertent decontextualization which can fulfill the function of platform filtering. Even while we will argue in the following sections that the mechanisms we identify provide evidence for intent, intent is also not strictly necessary. The process could proceed through a combination of the coordinated and the haphazard. In either case, the collage creator and the collage propagators are the decentralized actors primarily involved in the organization of bias (cf., Lukes, 1974) that underlies the disinformation campaign.

Discussion

Theoretical Implications

Our findings call attention to the issue of disinformation impediments, highlights the importance of what might be called the "throughput context" of disinformation spreading, and problematizes the assumption of passive facilitators in disinformation flows (cf., Marwick, 2018). The throughput context here becomes both part of the challenge facing a disinformation campaign (skeptical ancillary participants impeding disinformation) and components of the mechanisms we identify (how disinformation campaign actors may still leverage the characteristics of these contexts such as technological affordances and ambiguous situations). The mechanisms of evidence collaging and platform filtering leverage structural aspects of social media in order to overcome endogenous impediments to disinformation flows.

These characteristics of disinformation flows we identify are further complicated by the fact that disinformation can serve multiple functions in different cases, one of which is sometimes actually the creation of skepticism and mistrust. This function of disinformation is to perpetuate or exacerbate ambiguity by offering alternative narratives of events (Pomerantsev & Weiss, 2014; Wilson, Zhou, & Starbird, 2018) so that an uncertain environment emerges wherein individuals lose trust in authoritative sources of information because they are awash with conflicting information (Starbird, 2017). The skeptical remarks surrounding disinformation in the case we study, which disinformation agents in our case likely are trying to overcome, may in other cases actually be part of the intended effect of the disinformation campaign.

Disinformation and Intent

Definitions of disinformation tend to revolve around intent, which is often difficult to deduce when content is generated by an anonymous actor. Jack (2017) defines disinformation as "information that is deliberately false or misleading," while Benkler et al. (2018) build on Jack's definition and offer "manipulating and misleading people intentionally to achieve political ends." For observers, the primary challenge to identifying intent is distinguishing a user expressing their legitimate beliefs from the intentional spread of false content. Platform filtering, when combined with evidence collages, provides the basis for an argument for inferring intent from communication traces, and therefore for distinguishing disinformation from other forms of rumor spreading in social media environments.

In the initial thread about Joel V. that contains the evidence collage there are a substantial number of dissenting opinions and reported disconfirming pieces of evidence. Reasonable participants in that thread must therefore have had significant grounds for skepticism about the Joel V. rumor. However, the person who created the evidence collage aggregated only confirming evidence, and the collage was then removed from the conversational context where negative evidence was represented. The combination of these two factors implies that a person creating the collage was at least negligently ignoring disconfirming evidence and creating a collage that would only be compelling in an outside context where that negative evidence was not present. Our analysis of the tactics of disinformation in this case therefore provide additional evidence to infer intent on the part of the individuals posting the disinformation.

Decentralized Communication Power and Design Implications

We also consider the design of social media platforms as a factor in amplifying disinformation. In our observations, we saw that an important factor for the spread of the Joel V. disinformation campaign was the decontextualization of the content as it moved across conversations and platforms. Making certain platform design choices or setting protocols that facilitate tracking pieces of content that migrate across platforms could mitigate filtering effects. While the literature on communication power highlights the importance of gatekeeping, curation, and influencers to spread content, our case study shows that other sociotechnical features of platform design, such as decentralization, can advance disinformation campaigns, especially if the content is an image file.

The platform filtering effect we identified can be viewed as one way that disinformation agents take advantage of open and fragmented conversational contexts. Even if we observe a degree of rational discourse within an integrated conversational context, that discourse, which could lead to disinformation being debunked, can easily be discarded by reposting the distilled disinformation into a new context where no such dissent has taken place. In our case study, evidence collages pose a significant challenge for tracing origins as they are easily downloaded, edited, and re-uploaded by individuals repeatedly. Watermarking could mitigate this disinformation tactic by tagging images with a link to the conversational context through which it moves. These watermarks could be read in other parts of the same website or by other websites to either automatically link backwards or create a flag or warning. There is precedent for this kind of mechanism in Twitter's embedded tweets, but a watermarking mechanism could be more fully developed.

Conclusions

In our study we focused on the question of how disinformation spreads in the face of expressions of doubt or skepticism. Despite the existence of skepticism and other challenges to the disinformation narrative we studied, the disinformation campaign persisted. We attribute this persistence to a combination of the disinformation's identity-confirming characteristics, the selective curation of positive evidence into easily transportable evidence collages, and the mechanism of platform filtering. We argued that the latter two aspects indicate a willful intent to spread false content. We also discussed how these tactics take advantage of open and fragmented conversational contexts, and create a source of communication power for individuals with little inherent power deriving from their social positions. We point out that these tactics are also enabled by the structure of digital platforms, and suggest that watermarking image content could be one mechanism to partially counteract these tactics.

If the web and web platforms remain in their current form, disinformation campaigns will continue to scale. Not only have adversarial groups aligned and collaborated within specific web communities, they have also learned how to leverage the design of online communication infrastructure to reach new audiences. Previously, decentralization of the web was heralded as a value of web architecture. In recent years, however, this characteristic has been exploited as a tool by disinformation agents and media manipulators who use it to filter dissent and exploit computational propaganda techniques as they push content across platforms. Stakeholders including technologists, designers, regulators, researchers, and web users must challenge those with the power to change platforms to acknowledge how decentralized communication operates to serve both knowledge production (on the positive side) and disinformation (on the negative side); and must demand distributing the power to make design changes (cf., Frey, Krafft, & Keegan, 2019) so that commercial interests are not locking in the status quo. While there are many laudable values in technological design embracing decentralization, the current products of social media must encode accountability, transparency, justice, and co-design to overcome these new media manipulation tactics. The fate of our global techno-futures depends on it.

Acknowledgments

Special thanks to Matt Goerzen for curating key parts of our initial data sources and for insights about happenings at the UTR, and to Rishab Nithyanand, Becca Lewis, Kinjal Dave, Francesca Tripodi, Jeanna Matthews, Patrick Davison, danah boyd, Jeremy Blackburn, Brian Friedberg, Joshua Becker, and Wm Salt Hale for discussion, tips, and feedback. This work was supported in part by the Data & Society Research Institute, the Washington Research Foundation, and a Data Science Environments project award from the Gordon and Betty Moore Foundation (Award #2013-10-29) and the Alfred P. Sloan Foundation (Award #3835) to the University of Washington eScience Institute. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect those of the sponsors.

Notes

1. Remarkably, some on 4chan are explicit about their strategizing to engage in media manipulation through techniques such as evidence collaging. One poster advertises a "News Spam Kit" describing the strategy of making a narrative easily digestible and providing lists of news sources and Twitter accounts sympathetic to white supremacist or far-right narratives. The presence of the attitude behind this post within the 4chan/pol/community lends further credibility to the explicitly manipulative intentions behind collaging and propagandizing using collaging documented in our case study.

2. The total amount of relevant data consisted of six 4chan threads, four 8chan threads, one Discord channel, two threads from other far right forums (Zerohedge and FreeRepublic), an Everipedia post, six far-right blogs (Puppet String News, GotNews, The Gateway Pundit, Studio News Network, Freedom Daily, and YourNewsWire), one Google Group (alt.security.terrorism), eight eddit threads (from/r/The_Donald/,/r/conspiracy/,/r/uncensorship/, and/r/TheBlogFeed/, and/r/news/), screenshots of Joel V.'s Facebook page, and 15 tweets.

3. We have decided not to reprint the last name of the family in this article out of respect for their privacy.

4. 4chan threads are automatically archived by a site called 4plebs.org, while threads on related forums like 8chan are frequently archived by users using third-party websites, providing a lasting record of many popular threads that would otherwise disappear.

5. Evidence collaging and similar visual media styles have been components of several other instances of rumoring and media manipulation, including the notable cases of Pizzagate (Donovan & Friedberg, 2019) and the Boston Marathon Bombing Reddit campaign (see e.g., "Rumours can spread via social-media during crisis situations, but rumours can also create a crisis situation" https://rumoursandsocialmedia.wordpress.com/tag/boston-marathon/).

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the Alfred P. Sloan Foundation [3835]; Gordon and Betty Moore Foundation [2013-10-29]; Washington Research Foundation.

ORCID

P. M. Krafft http://orcid.org/0000-0001-8570-2180

References

Allport, G. W., & Postman, L. (1946). An analysis of rumor. *Public Opinion Quarterly*, *10*(4), 501–517. doi:10.1086/265813

Arif, A., Robinson, J. J., Stanek, S. A., Fichet, E. S., Townsend, P., Worku, Z., & Starbird, K. (2017). *A closer look at the self-correcting crowd: Examining corrections in online rumors*. In Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing. doi:10.1145/2998181.2998294

Barberá, P., Tucker, J., Guess, A., Vaccari, C., Siegel, A., Sanovich, S., & Nyhan, B. (2018). *Social media, political polarization*. and political disinformation: A review of the scientific literature. doi:10.2139/ssrn.3144139

Benkler, Y., Faris, R., & Roberts, H. (2018). *Network propaganda: Manipulation, disinformation, and radicalization in American politics*. New York, NY: Oxford University Press.

Bennett, L., & Pfetsch, B. (2018). Rethinking political communication in a time of disrupted public spheres. *Journal of Communication*, *68*(2), 243–253. doi:10.1093/joc/jqx017

Bennett, W. L., & Livingston, S. (2018). The disinformation order: Disruptive communication and the decline of democratic institutions. *European Journal of Communication*, *33*(2), 122–139. doi:10.1177/0267323118760317

Bordia, P., & DiFonzo, N. (2004). Problem solving in social interactions on the internet: Rumor as social cognition. *Social Psychology Quarterly*, *67*(1), 33–49. doi:10.1177/019027250406700105

Bozdag, E. (2013). Bias in algorithmic filtering and personalization. *Ethics and Information Technology*, *15*(3), 209–227. doi:10.1007/s10676-013-9321-6

Castellano, C., Fortunato, S., & Loreto, V. (2009). Statistical physics of social dynamics. *Reviews of Modern Physics*, *81*(2), 591. doi:10.1103/RevModPhys.81.591

Castells, M. (2012). *Networks of outrage and hope: Social movements in the internet age*. Malden, MA: Polity Press.

Castells, M. (2013). *Communication power*. New York, NY: Oxford University Press.

Chadwick, A. (2017). *The hybrid media system: Politics and power*. New York, NY: Oxford University Press.

Dailey, D., & Starbird, K. (2015). *It's raining dispersants: Collective sensemaking of complex information in crisis contexts*. Companion Proceedings of the 18th ACM Conference Companion on Computer Supported Cooperative Work & Social Computing. doi:10.1145/2998181.2998294

Dauber, C. E., & Winkler, C. K. (2014). Radical visual propaganda in the online environment: An introduction. In C. K. Winkler & C. E. Dauber (Eds.), *Visual propaganda and extremism in the online environment* (pp. 1-30). Carlisle, PA: Strategic Studies Institute and U.S. Army War College Press.

DiFonzo, N., & Bordia, P. (2007). *Rumor psychology: Social and organizational approaches*. Washington, DC: American Psychological Association.

Donovan, J. (2018). After the #Keyword: Eliciting, sustaining, and coordinating participation across the occupy movement. *Social Media + Society*, *4*(1). doi:10.1177/2056305117750720

Donovan, J., & Friedberg, B. (2019). *Source hacking: Media manipulation in practice*. New York, NY: Data & Society Research Institute.

Dutton, W. H., Reisdorf, B. C., Blank, G., & Dubois, E. (2019). Searching through filter bubbles, echo chambers. In M. Graham & W. H. Dutton (Eds.), *Society and the internet: How networks of information and communication are changing our lives* (pp. 228). New York, NY: Oxford University Press.

Frey, S., Krafft, P. M., & Keegan, B. (2019). *"This place does what it was built for": Designing digital institutions for participatory change*. The 22nd ACM Conference on Computer-Supported Cooperative Work and Social Computing (CSCW). doi:10.1145/3359134

Friedkin, N. E., & Johnsen, E. C. (2011). *Social influence network theory: A sociological examination of small group dynamics*. New York, NY: Cambridge University Press.

Friggeri, A., Adamic, L., Eckles, D., & Cheng, J. (2014). *Rumor cascades*. In Eighth International AAAI Conference on Weblogs and Social Media. Ann Arbor, MI.

Geiger, R. S., & Ribes, D. (2011). *Trace ethnography: Following coordination through documentary practices*. In 2011 44th Hawaii International Conference on System Sciences. doi:10.1109/HICSS.2011.455

Griffin, M. (2015). Visual communication. In G. Mazzoleni (Ed.), *The international encyclopedia of political communication* (pp. 1646). West Sussex, UK: John Wiley Sons.

Hawley, G. (2017). *Making sense of the alt-right*. New York, NY: Columbia University Press.

Heverin, T., & Zach, L. (2012). Use of microblogging for collective sense-making during violent crises: A study of three campus shootings. *Journal of the Association for Information Science and Technology*, *63*(1), 34–47.

Hine, G. E., Onaolapo, J., De Cristofaro, E., Kourtellis, N., Leontiadis, I., Samaras, R., & Blackburn, J. (2017). *Kek, cucks, and God emperor Trump: A measurement study of 4chan's politically incorrect forum and its effects on the web*. AAAI International Conference on Web and Social Media (ICWSM). Montreal, Canada.

Hodas, N., & Lerman, K. (2012). *How visibility and divided attention constrain social contagion*. In International Conference on Privacy, Security, Risk and Trust and International Conference on Social Computing. doi:10.1109/SocialCom-PASSAT.2012.129

Jack, C. (2017). *Lexicon of lies: Terms for problematic information*. New York, NY: Data & Society Research Institute.

Krafft, P., Zhou, K., Edwards, I., Starbird, K., & Spiro, E. (2017). *Centralized, parallel, and distributed information processing during collective sensemaking*. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. doi:10.1145/3025453.3026012

Leavitt, A., & Robinson, J. J. (2017). *The role of information visibility in network gatekeeping: Information aggregation on Reddit during crisis events*. In Proceedings of the ACM

Conference on Computer Supported Cooperative Work and Social Computing. doi:10.1145/2998181.2998299

Lerman, K., & Hogg, T. (2014). Leveraging position bias to improve peer recommendation. *PloS One*, *9*(6), e98914. doi:10.1371/journal.pone.0098914

Lukes, S. (1974). *Power: A radical view.* New York, NY: Macmillan.

Maddock, J., Starbird, K., Al-Hassani, H. J., Sandoval, D. E., Orand, M., & Mason, R. M. (2015). *Characterizing online rumoring behavior using multi-dimensional signatures*. In Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing. doi:10.1145/2675133.2675280

Marwick, A. (2018). *Why do people share fake news? A sociotechnical model of media effects*. Georgetown Law Technology Review, 474.

Marwick, A., & Lewis, R. (2017). *Media manipulation and disinformation online*. New York, NY: Data & Society Research Institute.

Mathison, S. (2009). Seeing is believing: The credibility of image-based research and evaluation. In J. Hughes (Ed.), *SAGE visual methods volume II: Documentation and representation* (pp. 31-44). London, UK: SAGE Publications.

Mendoza, M., Poblete, B., & Castillo, C. (2010). *Twitter under crisis: Can we trust what we RT?* In Proceedings of the First Workshop on Social Media Analytics. doi:10.1145/1964858.1964869

Meraz, S. (2009). Is there an elite hold? Traditional media to social media agenda setting influence in blog networks. *Journal of Computer-Mediated Communication*, *14*(3), 682–707. doi:10.1111/jcmc.2009.14.issue-3

Meraz, S., & Papacharissi, Z. (2013). Networked gatekeeping and networked framing on #Egypt. *The International Journal of Press/Politics*, *18*(2), 138–166. doi:10.1177/1940161212474472

Monterde, A., & Postill, J. (2014). Mobile ensembles: The uses of mobile phones for social protest by Spain’s indignados. In G. Goggin & L. Hjorth (Eds.), *The Routledge companion to mobile media* (pp. 429-438). New York, NY: Routledge.

Nadler, A., Crain, M., & Donovan, J. (2018). *Weaponizing the digital influence machine*. New York, NY: Data & Society Research Institute.

Phillips, W. (2015). *This is why we can’t have nice things: Mapping the relationship between online trolling and mainstream culture*. Cambridge, MA: MIT Press.

Phillips, W. (2018). *The oxygen of amplification: Better practices for reporting on extremists, antagonists, and manipulators*. New York, NY: Data & Society Research Institute.

Pomerantsev, P., & Weiss, M. (2014). *The menace of unreality: How the Kremlin weaponizes information, culture and money.* New York, NY: Institute of Modern Russia.

Ragin, C. C., & Becker, H. S. (1992). *What is a case? Exploring the foundations of social inquiry.* Cambridge, UK: Cambridge University Press.

Rosnow, R. L., & Kimmel, A. J. (2000). Rumor. In A. E. Kazdin (Ed.), *Encyclopedia of psychology* (Vol. 7, pp. 122–123). New York, NY: Oxford University Press.

Sayre, B., Bode, L., Shah, D., Wilcox, D., & Shah, C. (2010). Agenda setting in a digital age: Tracking attention to California Proposition 8 in social media, online news and conventional news. *Policy & Internet*, *2*(2), 7–32. doi:10.2202/1944-2866.1040

Shibutani, T. (1966). *Improvised news: A sociological study of rumor.* Indianapolis, IN: The Bobbs-Merril Company Inc.

Shin, J., Jian, L., Driscoll, K., & Bar, F. (2017). Political rumoring on Twitter during the 2012 US presidential election: Rumor diffusion and correction. *New Media & Society*, *19*(8), 1214–1235. doi:10.1177/1461444816634054

Shoemaker, P. J., & Vos, T. P. (2009). *Gatekeeping theory.* New York, NY: Routledge.

Simon, T., Goldberg, A., Leykin, D., & Adini, B. (2016). Kidnapping WhatsApp—Rumors during the search and rescue operation of three kidnapped youth. *Computers in Human Behavior*, *64*, 183–190.

Spiro, E. S., Fitzhugh, S., Sutton, J., Pierski, N., Greczek, M., & Butts, C. T. (2012). *Rumoring during extreme events: A case study of Deepwater Horizon 2010*. In Proceedings of the 4th Annual ACM Web Science Conference. doi:10.1145/2380718.2380754

Starbird, K. (2017). *Examining the alternative media ecosystem through the production of alternative narratives of mass shooting events on Twitter.* In Eleventh International AAAI Conference on Web and Social Media. Montreal, Canada.

Starbird, K. (2019). Disinformation's spread: Bots, trolls and all of us. *Nature*, *571*(7766), 449. doi:10.1038/d41586-019-02235-x

Starbird, K., Maddock, J., Orand, M., Achterman, P., & Mason, R. (2014). *Rumors, false flags, and digital vigilantes: Misinformation on Twitter after the 2013 Boston Marathon Bombing*. iConference 2014 Proceedings. doi:10.9776/14308

Thorson, K., & Wells, C. (2015). How gatekeeping still matters: Understanding media effects in an era of curated flows. In T. Vos & F. Heinderyckx (Eds.), *Gatekeeping in transition* (pp. 39–58). New York, NY: Routledge.

Tufekci, Z. (2014). Engineering the public: Big data, surveillance and computational politics. *First Monday*, *19*, 7. doi:10.5210/fm.v19i7.4901

Wall, T., & Mitew, T. (2018). Swarm networks and the design process of a distributed meme warfare campaign. *First Monday*, *23*, 5.

Wallace, J. (2018). Modelling contemporary gatekeeping: The rise of individuals, algorithms and platforms in digital news dissemination. *Digital Journalism*, *6*(3), 274–293.

Waltman, M. S. (2014). Teaching hate: The role of internet visual imagery in the radicalization of white ethno-terrorists in the United States. In C. K. Winkler & C. E. Dauber (Eds.), *Visual propaganda and extremism in the online environment* (pp. 83-104). Carlisle, PA: Strategic Studies Institute and U.S. Army War College Press.

Wendling, M. (2018). *Alt-right: From 4chan to the White House*. London, UK: Pluto Press.

Wilson, T., Zhou, K., & Starbird, K. (2018). *Assembling strategic narratives: Information operations as collaborative work within an online community.* Proceedings of the ACM on Human-Computer Interaction, 2(CSCW). doi:10.1145/3274452

Woolley, S., & Guilbeault, D. (2017). *Computational propaganda in the United States of America: Manufacturing consensus online*. Oxford, UK: Computational Propaganda Research Project.

Woolley, S., & Howard, P. (2017). Computational propaganda worldwide: Executive summary. Oxford, UK: *Computational Propaganda Project*.

Woolley, S. C., & Howard, P. (2016). Automation, algorithms, and politics | Political communication, computational propaganda, and autonomous agents — Introduction. *International Journal of Communication*, *10*, 9.

Woolley, S. C., & Howard, P. (Eds). (2018). *Computational propaganda: Political parties, politicians, and political manipulation on social media* (Reprint ed.). New York, NY: Oxford University Press.

Zannettou, S., Caulfield, T., De Cristofaro, E., Kourtellis, N., Leontiadis, I., Sirivianos, M., & Blackburn, J. (2017). *The web centipede: Understanding how web communities influence each other through the lens of mainstream and alternative news sources*. ACM Internet Measurement Conference (IMC). doi:10.1145/3131365.3131390