# Disinformation as Collaborative Work: Surfacing the Participatory Nature of Strategic Information Operations

**KATE STARBIRD,** Human Centered Design & Engineering, University of Washington
**AHMER ARIF,** Human Centered Design & Engineering, University of Washington
**TOM WILSON,** Human Centered Design & Engineering, University of Washington

In this paper, we argue that strategic information operations (e.g. disinformation, political propaganda, and other forms of online manipulation) are a critical concern for CSCW researchers, and that the CSCW community can provide vital insight into understanding how these operations function—by examining them as collaborative "work" within online crowds. First, we provide needed definitions and a framework for conceptualizing strategic information operations, highlighting related literatures and noting historical context. Next, we examine three case studies of online information operations using a sociotechnical lens that draws on CSCW theories and methods to account for the mutual shaping of technology, social structure, and human action. Through this lens, we contribute a more nuanced understanding of these operations (beyond "bots" and "trolls") and highlight a persistent challenge for researchers, platform designers, and policy makers—distinguishing between orchestrated, explicitly coordinated, information operations and the emergent, organic behaviors of an online crowd.

CCS Concepts: • **Human-centered computing → Collaborative and social computing • Social and professional topics → Computing / technology policy**

**KEYWORDS**

Social media; Information Operations; Disinformation; Media Manipulation

## 1 INTRODUCTION

In recent years, social media have been strategically leveraged by numerous and diverse actors for political gain—from the ongoing harassment of government-critical media outlets in the Philippines [59], to the manipulation of democratic processes in Britain [14] and the U.S. [58] in 2016, to the "coordinated inauthentic behavior" feeding recent tensions between India and Pakistan [29]. We've all been exposed to numerous terms that attempt to describe these growing problems (e.g. fake news, digital pollution [27], information disorder [100]. information war [68]) and give them historical context (active measures [9], disinformation [40], asymmetric warfare [91]). Here, we use the term *strategic information operations* to encompass efforts by individuals and groups, including state and non-state actors, to manipulate public opinion and change how people perceive events in the world by intentionally altering the information environment. These operations are a global phenomenon, with political, social, psychological, educational, and cybersecurity dimensions. Indeed, researchers across

This work is supported by the National Science Foundation (grants 1715078 and 1749815) and the Office of Naval Research (grants N000141712980 and N000141812012).

Author's addresses: kstarbi@uw.edu, ahmer@uw.edu, tomwi@uw.edu



https://doi.org/xxxxxxx

fields are convening to understand these operations and to craft solutions to address them—through policy, education, and technology design.

In this paper we argue that strategic information operations are a critical concern for CSCW researchers and that CSCW researchers can make important contributions to these conversations. Strategic information operations, and in particular *disinformation*, function to undermine the integrity of the information space and reduce human agency by overwhelming our capacity to make sense of information [66, 69]. They therefore strike at the core of our values. And they affect things we care about—for example, finding life-saving information during a crisis event [85, 95], organizing online for political change [87, 96, 104], and protecting online spaces from bullying and harassment [10, 97].

The online aspects of modern information operations renders them newly legible for examination—providing opportunities to study their motives, tactics, and effects through the digital traces they leave behind. The field of CSCW offers methods and theories that can help us analyze and make sense of these traces, particularly by focusing on the mutual shaping of the technological platforms, social structures, and human behavior. Here, we present three case studies of strategic information operations, each examined as a particular configuration of collaborative "work" within an online crowd. This perspective encourages us to focus on the participatory nature of these activities, moving on from reductive and exotified accounts of "bots" and "trolls" to more complex conceptualizations that account for how strategic information operations integrate into existing online communities and leverage those communities in dynamic ways to achieve their goals. Through this lens, we can attempt to account for aspects of these operations that are entirely "orchestrated" by a set of paid actors, as well as aspects that are more "organic"—i.e. emerging through implicit coordination among diverse actors who may not share the same goals as the operators or be aware of their role in the operation.

Our paper is organized as follows. First, we call out and define our terms—*strategic information operations* and its subtype, *disinformation*—providing historical context for each. Next, we present our framework for examining these operations as collaborative work, followed by a description of our methodological approach. We then present three recent, complementary case studies of information operations—Case 1: the participation of Russian trolls in politicized Twitter discourse leading up to the 2016 U.S. election; Case 2: a persistent, multi-platform campaign targeting the "White Helmets" in the Syrian Civil War; and Case 3: sustained conspiracy theorizing about crisis events that serves as a vector of political disinformation—to illustrate how our sociotechnical systems allow information operations to target, integrate with, and leverage the activities of online crowds. Finally, we discuss the implications of this intersection (between improvised citizen-based activity and orchestrated efforts) for those seeking to address the spread of problematically inaccurate information in online settings.

## 2 BACKGROUND

### 2.1 Strategic Information Operations Online

*Information operations* is a term social media companies like Facebook have adopted to describe organized communicative activities that attempt to circulate problematically inaccurate or deceptive information on their platforms [101]. These activities can be called "strategic" in reference to Habermas's theory of communicative rationality [33], which distinguishes between understanding-oriented communicative activities and strategic ones that are oriented towards other persons from a purpose-driven, utilitarian point of view (e.g. the calculative manipulation of others). Information operations as a term originates within the U.S. intelligence community, where it refers to actions of "soft warfare" that aim to degrade the decision-making capabilities of a geopolitical adversary by targeting human perception and cognition rather than physical infrastructures [44]. Certainly this term remains unsatisfying, for its abstract unspecificity, and for its connection to militaristic metaphors like "information war" that risk legitimizing nationalist or nativist sentiments. It also highlights particular actors and motives over others, inviting us to overlook—or flatten through simplistic equivalencies—the complexity and ambiguity inherent to many other closely related activities, such as advertising and public relation campaigns. However, it serves well enough as the uncertain terrain for the task of

broadening consideration of the work involved in these manipulation efforts, still mostly thought as something done to human crowds rather than *something human crowds do*.

The material affordances of social media platforms make them robust infrastructures for information operations [11, 12]. The tools and practices that have evolved to support the work of advertisers, such as effective impression management or segmenting and micro-targeting particular audiences [67], can also be leveraged by government and political operatives tasked with spreading certain narratives. Social media's crowd- and algorithm-driven information flows, and the easy sharing of content across large audiences, can also facilitate information operations by generating emergent and self-reinforcing effects. Finally, social media feeds combine different types of content such as news and entertainment, collapsing different contexts together [55]. This makes it difficult to discern the intentions behind any given piece of media content, especially given the challenges of establishing information provenance in these environments. Information operations can exploit this ambiguity by blending misleading rhetoric with accurate and inaccurate content as well as inaccurate sourcing information (e.g. by using fabricated identities). Jack has noted that by claiming that they are "setting the record straight" or confronting the status quo, information operations can disavow any intent to mislead and frame their agenda as educational or emancipatory in nature [40]. Journalists and social scientists can find themselves particularly disadvantaged in these circumstances—unless there is verifiable proof of intent to deceive, these groups risk reputational, professional and legal repercussions when investigating or making claims about information operations. Meanwhile, political operatives can exploit these professionals' cautiousness by using plausible deniability as a defense and spread misleading content without facing major threats to their own credibility [40].

## 2.2 Disinformation Operations

Some academics [21, 52, 59] and journalists [69] have theorized that not every information operation necessarily tries to foster support for a particular message or idea. Some operations instead focus on causing a distraction or creating uncertainty in ways that "kill the possibility of debate and a reality-based politics" [69]. In her *Lexicon of Lies*, Jack catalogs several such approaches [40]. For instance, the flooding of conversational spaces on social media with positive messages or attempts to change the subject has emerged to be a misdirectional tactic called *xuanchuan* that is associated with the "Fifty-Cent Army" in China [34, 47]. Within this paper we scope our consideration to *disinformation,* in part because our case studies prominently feature disinformation as a form of information operations.

Disinformation can be defined as information that is deliberately false or misleading [20]. However, it also has a more specific meaning, stemming from the etymologically related *dezinformatsiya,* a term with roots in Soviet intelligence operations. Bittman, a former practitioner of disinformation who defected to the U.S. in 1968 [74] and later became an academic researcher and teacher, describes disinformation as one form of Soviet *active measures*—offensive instruments of foreign policy that seek to extend influence and advance geopolitical goals by distorting the information environment and changing how people perceive the world. Active measures are akin to antithetical public relations—they seek to undermine, delegitimize, and denigrate a public image rather than working to improve it [9].

In the early 20th century Lenin anticipated that Western freedoms (e.g. of speech or the press) could be exploited for the purposes of subversion and the spread of propaganda [70] and he integrated these "informal penetration techniques" into policy [9]. But active measures as we describe them here emerged in the early 1960s when the role of the Soviet intelligence services shifted from passive intelligence gathering to the active dissemination of subversive messages [8]. The establishment of a dedicated apparatus to support this active role marked a "new era in which distortion concisely and purposefully taints the natural flow of information throughout the world" [8].

Soviet active measures were designed to have detrimental consequences on specific targets as well as wider-reaching geopolitical implications such as: undermining political, military or economic strength; creating or exacerbating existing rifts within societies; discrediting policies or representatives domestically and/or internationally; and influencing policy decisions. The complexity of these objectives dictated that active measures were deployed as part of a long-term strategy: it was acknowledged that a single active measure would have little impact, but when deployed persistently over a period of years the cumulative effects would be significant and long-lasting [9]. It is therefore useful to think of active measures generally, and disinformation specifically, not as a quality of a specific piece of information, but as a collection of information-actions—or a *campaign.* Additionally, evaluating

disinformation is less about the truth value of one or more pieces of information and more about how those pieces fit together to serve a particular purpose. To be effective, a disinformation campaign must be based around a "rational core" of plausible, verifiable information or common understanding that can be reshaped with disinformation—for example half-truths, exaggerations, or lies [9].

Journalists were often targeted as "unwitting agents" [9] in the spread of disinformation—for example through anonymous tips that offered a "scoop" or aligned with their existing beliefs. They would then unwittingly introduce the disinformation into the press without realizing the true intentions behind the transaction. It is reported that just a single sensationalist article with details of a conspiracy was often enough to generate interest from other outlets, facilitating dissemination and leading to shock among the public—with the hope that this would prompt action such as a protest [8]. Indeed there is evidence that conspiracy theories have been used, strategically, as a vector for political disinformation by exploiting existing rifts within society, for example the intentional spread of theories claiming that the CIA orchestrated JFK's assassination [52, 69] and that U.S. scientists created and spread AIDS as a biological weapon in Africa [19]. Historical accounts suggest that by crafting a message that appealed to existing political bias, groups at the extremes of society (e.g. the far left or far right) would be particularly amenable to conspiratorial content even if the source was questionable [9].

## 2.3 Contemporary Research of Online (Dis)Information Operations

Information operations that target social media have been growing in size and scope over the past decade. In their global inventory of organized social media manipulation, Bradshaw and Howard [11] report that since 2010, more than half a billion dollars have been spent by political parties and governments to research, develop, and conduct operations focused on manipulating public opinion over social media. These authors also found evidence in 2018 that these operations are now taking place in 48 countries [11]. Scholars have documented the effects these operations have had on political conversations in a range of contexts. In the Philippines, parties across the political spectrum have made use of hierarchized "click armies" that rely on digital workers—and very minimally on automated bots—to drown out dissenting opinions [59]. In Brazil, information operations harnessed large data sets of information about citizens held by corporations and governments to target different audiences during two Presidential campaigns, one Presidential impeachment campaign, and the election for the Mayor of Rio [4, 12]. In America, the National Intelligence Council has asserted that Russia engaged in significant efforts to disrupt "public faith in the democratic process" [58] during the 2016 presidential elections.

In these examples, and in existing contemporary research into strategic online information operations more broadly, the focus is often on explicitly coordinated activity—i.e. those activities conducted by automated "bots" [e.g. 1, 22, 103] or paid workers [e.g. 47, 59]. However, that perspective risks over-simplifying the dynamics of these operations, which often involve actors who are not explicitly coordinated and, in some cases, are not even aware of their role in the campaign—e.g. the "unwitting agents" described by Bittman [9].

## 2.4 Situating Strategic Information Operations as Collaborative Work

Online information operations are *participatory* in nature. Their messages spread through—and with the help of—online crowds and other information providers. Our work views strategic information operations online as collaborative work, a perspective that pushes us to expand our focus beyond "bots" and "trolls" to consider of the role of online crowds (unwitting and otherwise) in spreading disinformation and political propaganda. We adopt a CSCW perspective that provides a broad window for examining top-down, orchestrated work as well as other types of coordination, both explicit and implicit. It allows us to explore and account for both big "W" work that occurs within formal organizations [45] and little "w" work that emerges within distributed, online crowds [e.g. 57, 64, 71, 85]—as well as the intersection between the two.

The field of CSCW has also adopted and evolved sociotechnical theories—e.g. structuration [26], sociomateriality [61], and distributed cognition [39]—to assist in the investigation both of how technology shapes work within formal organizations [60, 61] and of how online environments facilitate and shape collective behavior in the "crowd" [82, 86, 95]. Through this sociotechnical lens, we can attempt to account for the mutual shaping of technological affordances, social structures, and human

action. Technological affordances are essentially what the platforms allow us (as users) to do—including features of the interfaces we use (e.g. to post or 'like' a tweet) as well as the algorithms that shape what we see of others and their content and how others see our content. Social structures include both the norms or rules that guide our actions within these systems, and the online communities that take shape around certain conversations. Human agency, i.e. the actions we choose to take, are shaped by these technological and social structures, but those actions also function to shape these structures in turn. For example, our networks of connections (our "social networks") are shaped by the technological affordances within these environments (we 'follow' someone and their information becomes part of our 'feed'). These networks of connections simultaneously enable and constrain certain interactions (we see the content of certain users and not others) that further shape our actions. Similarly, the technological infrastructure changes to adapt to our behaviors, sometimes gradually (through the addition of new features) and sometimes dynamically (we 'like' another user's post and the algorithm adapts to send us more of their content). Furthermore, actions we take over time contribute to the development of norms, guiding the behavior of other users within the environment. This view of the mutual shaping of technology, social structure, and human action allows us to surface some of the second- and third-order effects of strategic information operations in online environments.

Through this sociotechnical lens, we can attempt to address a critical challenge for researchers, platform designers, and policy makers—distinguishing between orchestrated, explicitly coordinated, information operations and the emergent, organic behaviors of an online crowd.

## 3 METHODOLOGICAL APPROACH

Perhaps obscured by the reactionary press coverage, online platforms like social media are likely only one element of multidimensional influence campaigns; however, online activities leave behind *digital traces* that provide new windows for studying these operations. Digital traces are records of online activities, left through the activities themselves, structured in a way that makes it possible to study them (i.e. to collect, store, and analyze), after the fact and at scale. Though these traces are often incomplete and imperfect, they do make the online components of modern information operations newly legible for researchers—allowing us to investigate the motives, tactics, and effects of strategic information operations in more public venues and at more rapid timescales. In other words, we no longer have to rely upon the accounts of defectors and the grey literature of other intelligence operators; we can see the operations for ourselves.

In investigating these traces, our research expands upon methodological innovation from *crisis informatics*, an interdisciplinary field with roots in CSCW, that examines how people use information and communication technology (ICT) to respond to crisis events such as natural and man-made disasters [63]. The methods of crisis informatics have evolved to enable the rapid collection, storage, and analysis of digital trace data created through the "mass participation" within online environments that occurs after disasters and other breaking news events. As the field of crisis informatics was taking shape, Palen and her colleagues argued that ICT-enabled interaction (and the resulting trace data) would allow us to examine human reactions to crisis events in new ways, even as it facilitated new responses and configurations of responders [38]. These same arguments can be applied to ICT-enabled information operations.

In recent years, we have employed and evolved the methods of crisis informatics [63] first for studying online rumors and misinformation [2, 53], and then later for studying intentional disinformation and information operations [e.g. 3, 84, 102] We employ a grounded, interpretivist, mixed-method approach that deeply integrates qualitative and quantitative analyses to provide insight into human behavior, as mediated by online platforms at scale. This methodological approach is informed by a sociotechnical perspective that guides us to examine how macro-level structures are shaped by micro-level interactions—and vice versa.

In our work, we repeatedly shift across perspectives, from high-level views (facilitated by visualizations and descriptive statistics) that help us identify patterns and anomalies, to close-up engagement with the content (qualitatively analyzing tweets, accounts, articles, etc.) to better understand the nature of those patterns and anomalies. Systematic content analysis also allows us to generate new hypotheses about underlying patterns in the trace data and conceive of new ways of exploring those patterns and testing those hypotheses (e.g. by creating a network graph with a unexplored edge property). As we iterate back and forth across different "levels", we generate, test, and

refine our interpretive explanations of the nature and relationships between structure, information dynamics, and human action.

In constructing knowledge from these trace data, our methodology adapts a grounded theory approach [15, 16] to include quantitative and visual representations as both artifacts for interpretative analysis and methods for stressing and refining our emerging theories. These methods are closely aligned with trace ethnography [25] and network ethnography [36], which provide similarly valuable lenses for following human activities through and across online platforms and noting how technologies and social structures (e.g. norms, social networks) shape online behavior.

### 3.1 A Brief Note on Reflexivity

Our approach acknowledges that it is impossible to remain "outside of" our study topic as researchers. As "human research instruments" [24] who are studying disinformation, the analyses we develop in this paper have been shaped by our practical, theoretical, experienced, and inexperienced lenses. Rather than treating these intersubjective elements as a reliability problem, we pay analytic attention to them to try and offer a richer understanding of disinformation as a complex phenomenon. To do this, we draw on qualitative traditions for fostering reflexivity in research design [7, 31, 49]. Some specific steps we take to foster reflexivity include: 1) involving multiple investigators from different backgrounds and 5+ countries in our research; 2) journaling about, reflecting on, and holding dialogs about how our work is shaping and being shaped by our epistemic beliefs, values, political perspectives and assumptions; and 3) briefly reporting on how these things may have come into play during the research process.

Crucially, we have observed that our own backgrounds and positions make it difficult to problematize certain aspects of information operations. For instance, when the information operations in our case studies amplified messages laden with progressive values (shared by members of our research team), we found ourselves wrestling with creeping doubt and skepticism about our interpretations of these operations as problematic—or as operations at all. We noticed these dissonances are intersectional. For example, messages that critique evidence using anti-imperialist or anti-positivist frames generated considerable tensions for those of us from decolonized countries and who are invested in the view that knowledge is situated and socially constructed. Surfacing these tensions helped us be generative and intimate in our research (because it reduces the distance between us and our "subjects"). These tensions sensitized us to how disinformation operations can effectively sow doubt and create confusion—not just in imaginary Others, but ourselves.

## 4 CASE STUDIES

Here, we present three case studies from our own research, each relying on long-term engagement with the specific context—spanning a year or more of intense study in each case and incorporating our original analyses with other researchers' and journalists' accounts of the phenomena. Though some of the empirical findings have been covered in prior work, the accounts presented here provide additional context from continued study, new details from supplementary data sources, previously unpublished insights, and new analyses guided by the sociotechnical perspective outlined above. Though similar in many ways, the three studies examine vastly different kinds of information operations—in terms of how they engage with, leverage, and shape the activities of the online communities within which they "work". The first case demonstrates a highly *orchestrated* campaign, the second shows *cultivation* of an online community, and the third reveals how *emergent* activity in an online crowd can resonate with the work of information operators. Individually, each case study provides a detailed and nuanced look at three persistent operations, all with ties to the Russian government's information apparatus—but also with evidence of other coordinated actors. Together, they provide complementary perspectices that reveal common elements and patterns of strategic information operations and highlight the complexity of disentangling organic online activity from orchestrated campaigns.

### 4.1 Case Study 1: Trolling Operations by the Internet Research Agency Targeting U.S. Political Discourse (2015-2016)

It is now widely recognized that the Internet Research Agency in St. Petersburg (RU-IRA) was conducting a years-long information operation on social media that was leveraged, in part, to influence political views in the United States leading up to the 2016 election. Evidence supporting this view has been reported in academic research [41], investigative journalism [65], intelligence committee reports [58], and by the platforms themselves [92, 93, 101].

Our lab initially encountered these operations accidentally, while studying other online phenomena—in particular, online discourse about the #BlackLivesMatter movement. In November 2017, we published a paper [87] about that highly polarized discourse on Twitter, examining "framing contests" between politically left-leaning, pro-#BlackLivesMatter accounts and politically right-leaning, anti-#BlackLivesMatter accounts. Shortly after publication, the U.S. House of Representatives Intelligence Committee released a list of Twitter accounts that had been determined to have been operated by the RU-IRA [94]. Upon first seeing that list of accounts, we recognized several from our study, including some that we had featured in our paper. After systematically cross-checking those accounts against our #BlackLivesMatter data, we found that RU-IRA accounts were embedded in, and in some cases quite influential within, both "sides" of that polarized conversation (see Figure 1). Later, Twitter released a full data set of all of these accounts and all of their tweets, and we were able to see how the conversation we had studied (#BlackLivesMatter) fit within the broader operations of the RU-IRA. As it turned out, we had stumbled into a significant element of their operation.

*4.1.1 Data and Methods:* This case study incorporates four different stages of analysis, each based on different data. It begins with our initial #BlackLivesMatter study [87], relying on data collected from Twitter related to shooting events in 2016. It then shifts to focus specifically upon the role of RU-IRA accounts in that same dataset [3], and then pulls back to examine the online activities of those accounts more broadly through qualitative analysis based on available trace data, accessed through the Internet Archive [90]. Finally, it makes use of the more recently released archive of RU-IRA operations [93] to add important context to earlier findings by examining the entire scope of those operations. Though the specific analyses varied across the different parts of this study, each was informed by our broader methodological approach, which deeply integrates qualitative and quantitative (including visual) methods to provide a grounded, interpretative explanation of the phenomena.

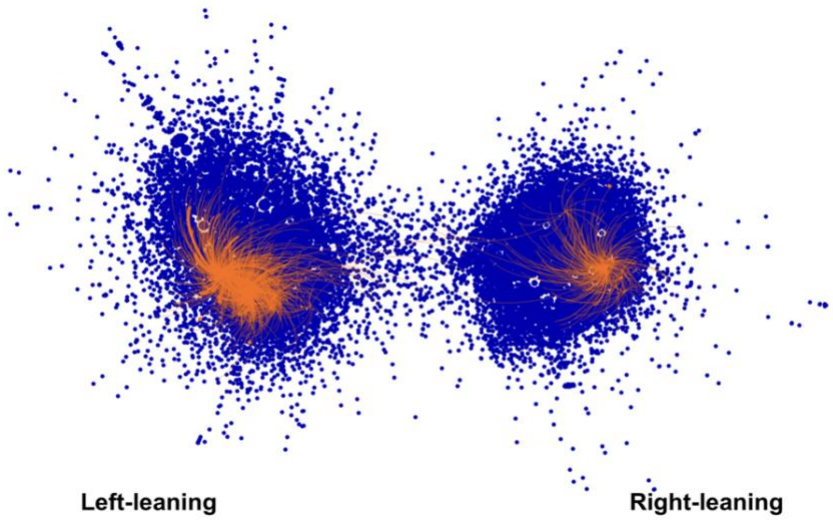


**Fig. 1. Retweet Network Graph: RU-IRA Agents in #BlackLivesMatter Discourse. The graph (originally published [3]) shows accounts active in Twitter conversations about #BlackLivesMatter and shooting events in 2016. Each node is an account. Accounts are closer together when one account retweeted another account. The structural graph shows two distinct communities (pro-BlackLivesMatter on the left; anti-BlackLivesMatter on the right). Accounts colored orange were determined by Twitter to have been operated by Russia's Internet Research Agency. Orange lines represent retweets of those account, showing how their**

**content echoed across the different communities.**
**The graph shows IRA agents active in both "sides" of that discourse.**

*4.1.2 Synthesized Findings: RU-IRA Agents Targeted, Infiltrated, and Cultivated Politically Active Communities Online:* RU-IRA agents "worked" together through the operation of more than 3000 accounts that presented themselves as people and organizations. About half were active in English-language discourse. The others primarily targeted Russian-speaking audiences. Focusing on the English-language accounts, using content and network analysis, we identified three distinct types of accounts: 1) "local" news accounts that reposted headlines from actual news in those markets; 2) "hashtag warrior" accounts that initiated and spread humorous hashtags in a game-like fashion to gain attention and followers; and 3) highly personalized accounts enacting politically active U.S. citizens. Among the third type, there were two distinct groups, one with accounts that impersonated African Americans and #BlackLivesMatter activists, and another that impersonated white, politically active conservatives, primarily from southern U.S. states.

RU-IRA accounts impersonating members of these two politically active groups—the African American "left" and the white, conservative "right"—created about one-third (more than 1M tweets) of all the English-language tweets posted by RU-IRA accounts. More importantly, this rather small subset of RU-IRA accounts (~300) garnered 85% of all English-language retweets (18.5M retweets). In other words, accounts impersonating activists within these groups were about a third of the RU-IRA's overall English-language operations, and by far their most successful in terms of receiving traction in the online crowd. Content analysis suggests that these accounts operated towards multiple objectives including: amplifying political divisions in the U.S.; criticizing and demotivating support for U.S. presidential candidate Hillary Clinton on the left; and promoting candidate and later President Trump on the right.


**Fig. 2. Example #BlackLivesMatter related content that was circulated by RU-IRA social media accounts to different audiences.**

There is now a significant body of evidence documenting how RU-IRA agents targeted the African American community, in part by impersonating #BlackLivesMatter activists [3, 17, 75]. Our analyses reveal that they also specifically targeted, and had some success infiltrating, "grassroots" conservative activist communities (see Figure 2 for examples). We quote "grassroots" here, because some of the communities have themselves been shaped by domestic campaigns to build conservative networks on Twitter. A notable example is the Patriotic Journalist Network (PJNET), a group of Twitter activists that use coordinated—and in some cases automated—tweeting practices to simulate or trigger virality for conservative messages [87]. PJNET is an online organization that has been active since at least 2014. The group has a suite of online tools that encourages on-message tweeting, especially of graphical memes. Group members use the #PJNET hashtag in their profiles and tweets, which facilitates

recruitment into the group and encourages "follow-back" practices among group members. These practices have affected the network structure within the right-leaning community—an effect documented by Stewart et al. [87] in #BlackLivesMatter discourse and by Supovitz et al. [89] in a study on online activism around #CommonCore. The IRA purposefully targeted the PJNET community, using their hashtags and cultivating their influencers. The #PJNET hashtag was among the top-30 most retweeted terms by RU-IRA accounts—tweeted more than 10,000 times. Most of these were retweets of non-IRA accounts who appended #PJNET to their tweets. The RU-IRA repeatedly amplified influencers in the PJNET community, including one account they retweeted more than 1140 times and who, in turn, followed at least seven different RU-IRA accounts.

To infiltrate such communities, RU-IRA agents created consistent, persistent, political personas that reflected caricatures of U.S. political participants. These personas were formed, in part, through mimicking the online presentations and behaviors of real U.S. citizens. In the full RU-IRA data, we can see an evolution of the accounts as they become more sophisticated—behaving more and more like "real" people online, and specifically converging with the presentation of highly political online accounts. A sociotechnical perspective suggests that the activities of RU-IRA accounts would also have shaped the social structures—i.e. the social norms and network ties—of the online communities they targeted, and consequently the actions of other users in those communities. This hypothesis is reflected in claims that the RU-IRA campaign was, in part, designed to "sow discord" [58, 69]. In other words, the RU-IRA operations targeted, infiltrated, and cultivated politicized online communities.

Journalistic accounts based on the account of activist Lyudmila Savchuk, who worked undercover at the Internet Research Agency's facility in St. Petersburg [56, 78], describe the "big-W" work of RU-IRA agents as taking place within that facility and organized in a top-down fashion, with supervisors providing lists of topics and agents focusing their "troll" accounts on those topics for their shifts. Though this work was managed by the formal organization, individual workers were empowered to create their own personas and improvise as they created content aligned with the organization's objectives. From this view, the work of RU-IRA agents can be classified as clearly orchestrated.

Though our analyses in some ways confirm this view of the RU-IRA operations as orchestrated, they also reveal a concerted strategy to integrate into organic online communities. RU-IRA agents did this in a variety of ways—by impersonating activists within those online communities, building networks within those communities through interactions enabled by the platform (e.g. liking, retweeting, following), and at times directly contacting other "real" activists or influencers within the community. There is evidence that, in a few rare cases, RU-IRA agents directly collaborated with activists, e.g. to organize physical protests in the U.S. [98]. In these ways, both indirectly and directly, unwitting activists in the U.S. began to coordinate their actions with RU-IRA agents.

However, it is important to stress that the online communities themselves were not products of those operations, but instead were functioning as unwitting hosts. Underscoring that point, the first RU-IRA tweets about #BlackLivesMatter do not appear until about 18 months after the movement began. This supports an understanding that, though they participated in the conversation and made some progress towards integrating into that community, the RU-IRA were not part of the foundational stages of the #BlackLivesMatter movement. Instead, they targeted an existing organic movement for infiltration and cultivation. Though their information sharing and engagement behaviors often aligned with that movement's ethos, their objectives were markedly different, as they sought to guide the community taking part in that movement towards political views favorable to Russia's long-term goals. The RU-IRA conducted parallel operations within the conservative, pro-Trump online community. On that "side" of the conversation, we can also see them targeting less organic movements, piggybacking off of existing domestic influence operations, like #PJNET, whose roots also appear to predate RU-IRA infiltration.

## 4.2 Case Study 2: The Disinformation Campaign Targeting the White Helmets

This second case study examines strategic information operations in the context of armed conflict—the Syrian Civil War. In particular, we focus on the disinformation campaign targeting the White Helmets, a volunteer humanitarian response group that works in rebel held areas of Syria. The White Helmets provide search and rescue assistance and medical aid to people affected by the conflict—primarily victims of air strikes perpetrated by the Syrian government and their Russian allies. The group also documents their work, publishing videos showing the human impacts of the conflict. In 2016, these efforts helped to garner international attention to and sympathy for Syrian citizens suffering at the

hands of their government—attention that might have moved the international community to call for action against President Assad and his allies. However, the group also became the target of a multi-party online campaign to delegitimize them [50, 77]. They were labelled tools of foreign influence, called criminals and terrorists, and accused of staging chemical weapons attacks and other events (see Figure 3 for an example). To a significant extent, this campaign worked. If you go online and search for the White Helmets, you will likely be overwhelmed by the critical content meant to delegitimize them, silence their voices, and justify targeting them for violence (in violation of international law). This multi-dimensional disinformation campaign reflects a complex, strategic information operation that includes both orchestrated and organic elements.

*4.2.1 Data and Methods:* We have been studying this conversation since June of 2017. Our data include an ongoing collection of Twitter data related to the Syrian conflict, scoped to tweets that contain an explicit reference to the White Helmets. Again, our work in this context applies mixed methods—qualitative, quantitative, and visual—in a grounded, interpretative approach. We examine the data at various "levels", shifting from micro- to meso- to macro- scales (and back again) to understand the content (competing narratives), structure (social networks) and dynamics (information flows, interactions) of the conversation [102]. We conduct content analysis of tweets and accounts, examine temporal patterns of content production (to look for things like coordinated action), and create network graphs to look at relationships between accounts (to understand the role of social networks and communities in the production of disinformation). We also use the links within tweets to look beyond Twitter at the surrounding information ecosystems that contribute to the White Helmets discourse [84]. Most recently, we have been conducting a cross-platform analysis to look at the role of YouTube in supporting this discourse.


**Fig. 3. Sample tweet amplifying an anti-White Helmets narrative.**

*4.2.2 Synthesized Findings: Cultivating an Online Activist Community; Filling Data Voids with Geopolitical Propaganda:* Focusing first on the Twitter conversation, one major finding is simply that anti-White Helmets voices overwhelm and drown out pro-White Helmets voices. When we look at the social structure of the White Helmets discourse, we find two distinct communities of accounts—one that

produced content in support of the White Helmets and another that produced content criticizing the White Helmets. Interestingly, critical accounts outnumbered pro-White Helmets accounts (by about 40%). Even more interestingly, those critical accounts were more consistently active and produced about three times as much Twitter content as the pro-White Helmets accounts. That critical content reflected many of the narratives we described above—e.g. claiming the White Helmets associated with terrorists and accusing them of various crimes. Importantly, the accounts that constituted the community of anti-White Helmets voices, especially the core accounts that produced the majority of content in that discourse, were not automated/bot accounts, nor do they appear to be predominantly paid troll accounts. Instead, most seem to be authentic, online "information activists" who devote significant personal resources to tweeting in support of this explicitly anti-White Helmets political agenda. Many have bridged from other related causes—including "anti-war" activism and support of the Palestinian cause. A few prominent accounts in the community appear to be agents of foreign governments or other groups (e.g. Hezbollah) who blend into these online activist communities, although it is very difficult to distinguish between authentic activists and those imposters. Some of the most influential accounts turn out to be Western "journalists," who rose to prominence through their anti-White Helmets content production (through tweets, blogs, and articles)—in part due to cross-platform amplification from state-sponsored media outlets (for example, broadcast interviews on Russia's flagship media outlet, RT). The tweet in Figure 3 was sent from the Twitter account associated with RT's UK current affairs program, which features an interview with one of the prominent journalists in the anti-WH conversation and demonstrates this cross-platform promotion.

Within the boundaries of Twitter, we can see ambient "work" by the activists, journalists, media outlets, and agents to amplify each other's content (through likes and retweets), as well as more explicit coordination work across these diverse actors. In one common tactic, an account will call attention to a piece of content (e.g. a tweet supporting the White Helmets) from other members of the group by adding a list of @mentions to a tweet quoting or linking to that content. Here's a template of the structure of one of these call-out tweets:

```
@activist1: @journalist1 @activist2 @activist3 @journalist2 @activist4
@agent @activist5 Hey, check this out! White Helmets are at it again.

<quoted tweet with a statement supporting the White Helmets or
portraying them in a positive light>
```

These mass-mention tweets result in a cascade of reactions, as the mentioned accounts mobilize to amplify their group's message (e.g. retweet @activist1's tweet) and to explicitly challenge (e.g. through replies to the original post) the pro-White Helmets claim and/or its author. The effect on the account whose tweet was quoted/replied-to can be experienced as a form of "dog-piling" [43]. Over time, these efforts function both to drown out and drive out (through intimidation and harassment) content supporting the White Helmets. Though the initiating account never explicitly says, "let's all go tweet a reply challenging the quoted tweet", this activity follows one of a few set routines that are repeatedly enacted and can be easily deciphered, copied, and learned by new members of the community. In this way, the "work" of this online community is loosely coordinated through a shared set of practices and routines. Although there are accounts within this community that are affiliated with influence operations from multiple state-actors, there is little evidence of explicit, top-down orchestration by those actors. However, it is important to note that it is possible and even likely that other forms of coordination are occurring within this assemblage of actors that are not visible to researchers, for example through direct messages on Twitter or other communication platforms, and actors that may appear to be authentic activists but who are in reality paid agents. This highlights the challenge of unraveling the *organic* from the *orchestrated* that we will return to.

Looking beyond Twitter to the surrounding media ecosystem (using URL links within tweets), reveals a similar asymmetry in the content and structure of the White Helmets' discourse—i.e. anti-White Helmets articles are cited far more often than pro-White Helmets articles. Through closer analysis (documented in detail in previous work [84]), we discovered a tightly connected network of websites that repeatedly share content, via copy-pasted articles (often word for word) criticizing the White Helmets (see Figure 4). We termed this content-sharing network a media "echo-system" and its structure reflects several different dynamics, including explicit content-sharing relationships between

websites, opportunistic appropriation of articles outside this network, and the use of tools that allow for easy republishing of content across certain publishing platforms.

Though it is difficult to attribute any of these effects to explicit coordination, patterns of activity suggest that the structure and dynamics of this echo-system reflect both organic properties and targeted influence campaigns. Among the most influential websites in this echo-system are a small number of self-described "independent" media outlets and think tanks that consistently share messages aligned with Russian and, in some cases, Iranian government interests. Other influential websites are media outlets explicitly funded and/or heavily influenced by the Russian (RT, Sputnik), Iranian (Farsnews), and Syrian (syrianews.cc) governments. These two types of prominent websites produce the majority of original anti-White Helmets content, which is then copied-and-pasted across a larger number of what appear to be diverse media outlets with varying ideological orientations—from TruePatriot to ActivistPost to VeteransToday to TheFreeThoughtProject. Many of these websites appear to be ideologically and/or financially motivated. It is also possible that the list includes imposter sites run by certain state and even non-state actors, a strategy that was documented in the RU-IRA activities described above. A closer look at this echo-system reveals a kind of micro-targeted strategy, where the same anti-White Helmets content is presented within different wrappers (different websites) that appeal to different audiences. Reflecting something about the shared infrastructure of disinformation, there is significant overlap with the network graph of domains supporting conspiracy theorizing about crisis events (see Figures 4 & 5).

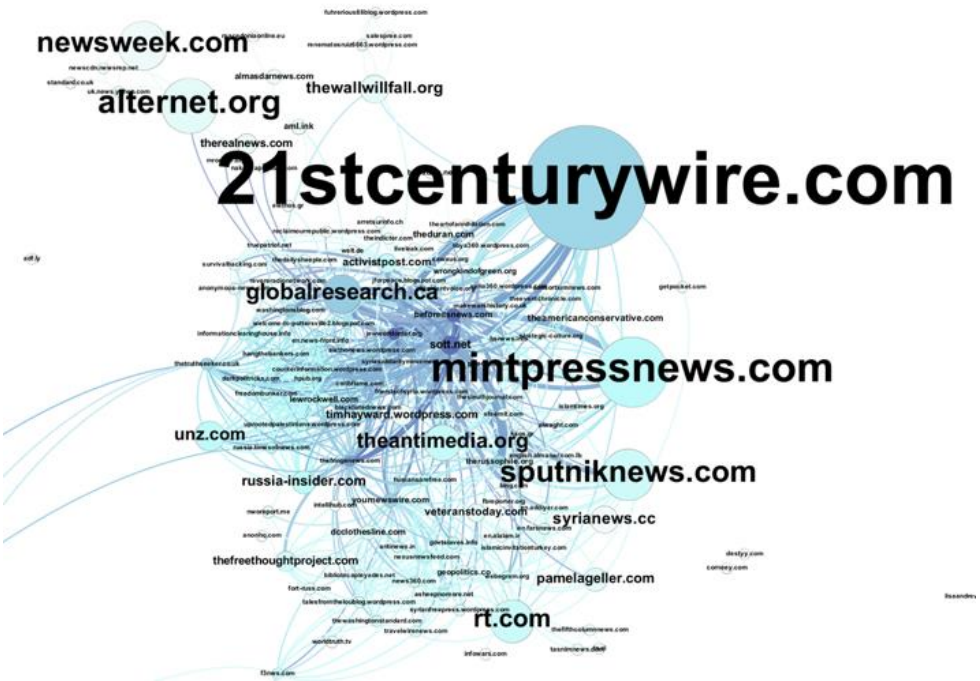


**Fig. 4. Domain Network Graph: Content Sharing across the "Alternative" Media Echo-System. Each node is a website that hosted an article cited in the White Helmets tweets. View is focused on domains that shared anti-White Helmets content. Nodes are connected when the same article appeared on both domains. Nodes are sized by the number of tweets citing that domain.**

Examining these structures and activities from a tactical point of view suggests that information operations—in particular disinformation campaigns—use multi-dimensional, multi-channel methods to flood and pollute information spaces. We can see this tactic applied across social media platforms in the anti-White Helmets campaign. Voices critical of the White Helmets dominate content-production on

several social media platforms, including Twitter and YouTube, where a relatively small number of highly active accounts consistently produce, interact with, and amplify anti-White Helmets content—and mobilize explicitly to challenge and drown out pro-White Helmets content. But the tactics aren't limited to social media; they extend to the surrounding media ecosystem where a relatively small amount of original content is re-shared and repurposed across a large number of websites, creating an echo-effect. These tactics can act to "game the system," pushing these tweets, videos, and articles up into the top search results for "White Helmets", essentially hijacking the conversation. This activity functions on one level to simply spread preferred narratives, challenging and delegitimizing the group. Indeed, the anti-White Helmets campaign seemed to be effective at overwhelming the White Helmets' efforts to draw garner sympathy and solidarity from Western audiences and at making them targets for political violence and retribution. At another level, this activity works to pollute the system and undermine trust in the information space more broadly.

Our analysis reveals the anti-White Helmets campaign to be a persistent, cross-platform, strategic information operation that was at least partially effective in their aims. This campaign had significant support from the Russian disinformation apparatus, including several of their media outlets (RT, Sputnik, In the NOW). There is also evidence of other state (Iranian) and non-state (Hezbollah) actors. However, this campaign is significantly different from the RU-IRA operations described above. Though there are some orchestrated elements, a significant portion of this activity appears to be organic—with considerable interplay and resonance between the two. From a sociotechnical perspective, these geopolitical entities are integrated into the sociotechnical infrastructure of this media ecosystem, and they shape the system in multiple ways, for example by hosting content on flagship media outlets and by pushing content out through ideologically-aligned, "independent" websites. They also actively support friendly journalists, publishing their content, amplifying their voices, and enhancing their professional reputations. So this campaign does reflect some elements of orchestration. But there are also many sincere actors or "unwitting agents" [9] in this space, online activists whose views have been shaped by the information operation (and whose actions are further shaped by the social and technical structure of these media spaces), and who actively support the generation and dissemination of the operation's preferred narratives. In many cases these agents are "journalists" whose reputation is so closely tied to their output of disinformation-aligned content that it is difficult to assign a single motivation to their participation. Considered as part of an information operation, this activity is perhaps best described as *cultivated* rather than orchestrated.

### 4.3 Case Study 3: The Online Ecosystem Supporting Conspiracy Theorizing about Crisis Events

This final case study explores how conspiracy theories of crisis events take shape on social media and with the support of an "alternative media" ecosystem that feeds and shapes them. Over and over again, and with increasing frequency in recent years, man-made crisis events such as school shootings and terrorist attacks have sparked online conspiracy theorizing (like the tweet in Figure 5 below)—claiming that the event did not happen in the way that media and government officials are portraying it.


**Fig. 5. Example tweet spreading a "false flag" narrative.**

As massive numbers of people converge online to make sense of a tragic event, a subset of that participatory audience begins to build and share theories that the event was a "hoax" staged by "crisis actors" or a "false flag" where the widely recognized suspects are not the real perpetrators, but are being framed by a secret group of powerful actors. On social media platforms, we can see these theories evolve over time, as the audience assembles evidence to support their theories and negotiates to find the theory that fits the available evidence best. Articles on "alternative news" websites, such as Alex Jones' InfoWars, often add fuel for the theories and "evidence" for these discussions. Though the explanations are dynamic and adapted to fit each particular event, the theories are consistent in that they begin with the idea that the narrative being promoted by mainstream media and government officials is false, and they work to find another "alternative narrative".

Our research team has accidentally been studying this behavior since 2013, when we initiated a research project focused specifically on the spread of online rumors during crisis events. And in early 2016, we began to see how this conspiracy theorizing of crisis events was connected to politically-motivated disinformation.

*4.3.1 Data and Methods:* Building from that early work, this case study includes insights from analyzing numerous "crisis actors" and "false flag" rumors across several distinct events between 2013 and 2016, including the Boston Marathon Bombings, the San Bernardino shootings, the Umpqua School shootings, and the Paris Attacks. Our research team identified, scoped, and in some cases did tweet-by-tweet coding of these "rumors"—reading tweets to understand how the rumors took shape, evolved, and were challenged on Twitter. Some of these analyses appear in published work [e.g. 3, 53].

This case study also features a multi-part analysis of conspiracy theorizing that took place on Twitter in the aftermath of shooting events in 2016 and an investigation of the online websites that supported those efforts [83]. The seed data for this work was a Twitter collection of shooting-related terms (e.g. shooting, shooter, gunman, etc.) that was active for nine months (January to October) in 2016. Using links in those tweets and co-sharing patterns created by users, we created a "domain network graph" (see Figure 6) that revealed some of the structure of the media ecosystem that surrounded and supported this conversation. Next, we coded the different web domains according to whether they contained content supporting the conspiracy theories of those shooting events (Figure 6, red), explicitly challenged those conspiracy theories (Figure 6, blue), or just got pulled into the conversation for being a "straight" news story about the event (Figure 6, yellow). Though yellow domains did not explicitly share a conspiracy theory in our dataset, yellow domains that are integrated into the red section of the graph are ones that conspiracy theorists often turn to for their news. Finally, we conducted an in-depth, qualitative content analysis of the different web domains, not limited to content about conspiracy theories of crisis events, but examining their content more broadly as well as their about pages and mission statements. The web domain content analysis was completed in December 2016.

*4.3.2 Synthesized Findings: A View Down the Rabbit Hole at the Disinformation Ecosystem:* Analysis of the Twitter data reveals a group of online conspiracy theorists, working collectively to produce, evolve, and amplify their theories with the support of an "alternative" media ecosystem. Their conspiracy theorizing about crisis events is not controlled by an outside actor or group. Instead, it emerges "organically" through the activities of group members who have come to share an epistemology that views world events as controlled by powerful, sinister actors and "mainstream" media as co-conspirators that help to hide that truth. This group does not wait for an external disinformation agent to give them the signal to start theorizing. Instead, each new event becomes interpreted through the same lens, and becomes a catalyst for another round of conspiracy theorizing. The participants work together to assemble available evidence to support their theories, often adjusting their alternative narrative to accommodate new pieces of information. At times, the group argues and negotiates to determine which direction the conspiracy theory should take—e.g. after the Orlando Pulse nightclub shooting in 2016, several accounts went back and forth debating whether the event was a "hoax" (i.e. it didn't really happen) or a "false flag" (i.e. it happened, but not in the way the media portrayed), eventually settling on the latter narrative as a better fit with emerging evidence.

This crowd of crisis event skeptics appears to be largely constituted by true believers (though there is evidence that some participants have other motivations—both financial and political). The online nature of their activity allows for an epistemic community to take shape around these conversations, through following relationships on Twitter, and in places like r/conspiracy on reddit. It also allows for

other people converging online to make sense of a crisis event to be passively exposed to this activity—as well as the epistemology that drives it and the techniques used to promote it. Thus, this activity can act to recruit others into the community and teach them the social rules of participating there. Studying this behavior allows us to see the intersection of platform features, social and socio-technical structures (network ties, online norms), and human behaviors that constitute the phenomenon of going "down the rabbit hole".

Our structural graph of the media ecosystem supporting this conversation (Figure 6) reveals the theorizing to be supported by a mix of clickbait news, self-described "alternative" and "independent" media, as well as government-affiliated domains such as RT and SputnikNews (Russia) and PressTV (Iran) and geopolitical think tanks associated with specifically pro-Russian stances. These websites do not initiate each episode of conspiracy theorizing, but their articles do shape, amplify, and sustain those conversations. Web articles function to assemble the different pieces of evidence into a single, coherent narrative—an "alternative narrative" to compete with the one that appears on "mainstream" media outlets. These articles are then repeatedly posted to social media through share buttons, and many get re-shared across other web domains, expanding the audience and exposure. The website content is also less ephemeral than the social media activity, helping to sustain the conspiracy theory over time.

Another significant finding (first reported in [83]), revealed when we look more deeply at the broader content on the web domains in red (Figure 6), is that many of these websites host content supporting not just one conspiracy theory, but many. Most of the websites in red host dozens of articles promoting different conspiracy theories across seemingly disparate topics and domains—e.g. climate change denial, anti-vaccine pseudoscience, claims that Obama is really a lizard-like alien and that Hillary Clinton runs a pedophile ring under a pizza place in Washington DC, theories that a shadowy cabal of rich Jewish men secretly run the world, 9-11 trutherism, flat earth theory, etc. These are very different theories, but they often reflect a shared epistemology—one that pushes people to question science, journalism, and the integrity of democratic institutions. Relatedly, much of the content on these domains was (as of December 2016) overtly political, but their politics were not traditionally "right" or "left". Instead, they promoted nationalist and populist messages and many were consistently supportive of U.S. President elect Trump, Russia's President Putin, and the geopolitical aims of Russia.

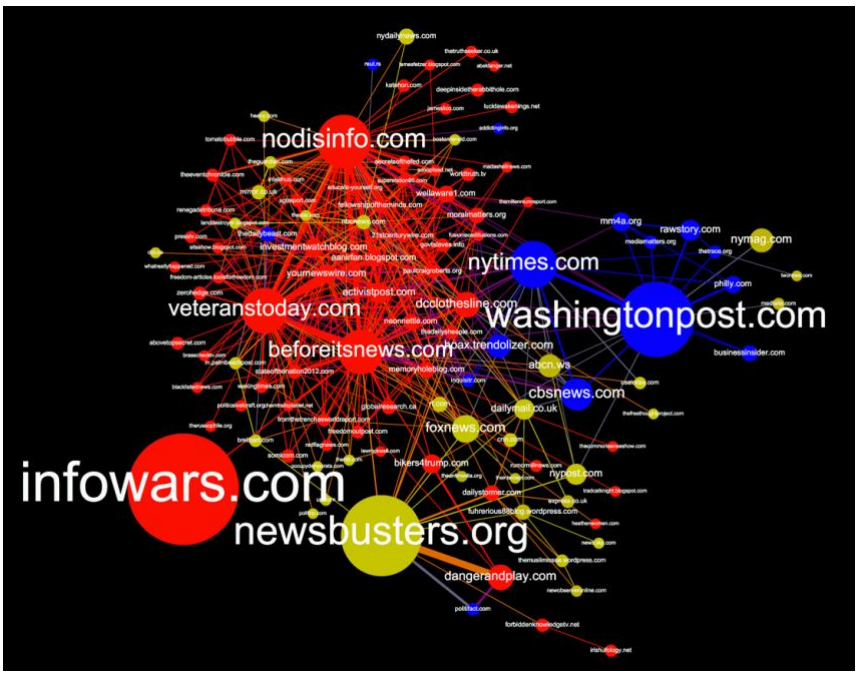


**Fig. 6. Domain Network Graph of Media Ecosystem that Surrounding Conspiracy Theorizing about Shooting Events in 2016. Nodes are domains cited in tweets referencing "false flag" or "crisis actors" conspiracy theories of shooting events. Nodes are sized by the number of tweets**

**linking to those domains. Edges are created when the same user sent tweets linking to both domains. Edges are weighted by the number of different users who tweeted linking to both domains. Red nodes are domains that hosted content explicitly promoting one or more conspiracy theories of the shooting events. Blue domains explicitly challenged those conspiracy theories. Yellow domains neither explicitly promoted nor explicitly challenged the conspiracy theories of those shooting events, but were pulled into the conversation for being a "straight news" article about the shooting event.**

This convergence of conspiracy theories and populist political messages likely has multiple causes and diverse effects. There are technical or structural factors in this convergence, whereby the tools of publication and content-sharing agreements in the alternative media ecosystem encourage low-budget websites to fill their pages with this free or low-cost content. There may also be strategic reasons for information operators to push this kind of content across these media ecosystems. It has been argued that a person who believes one conspiracy theory may be more susceptible to another [88], and so these websites could be acting purposefully as gateways from one conspiracy theory to another. There are echoes of other Soviet-era conspiracy theories in this ecosystem as well, demonstrating how the effects of information operations can be diffuse and long-lasting.

One way to view the online conspiracy theorizing described in this case study is as a largely emergent and self-sustaining activity that is shaped both by current sociotechnical systems and social structures as well as by lingering norms and epistemologies that have, in turn, been shaped in subtle ways, and to some extent over decades, by strategic information operations.

## 5 DISCUSSION

In this paper we have conceptualized information operations as sociotechnical phenomena that rely on a variety of actors and structures to successfully disseminate problematic kinds of information, and we have presented three case studies that demonstrate different kinds of information operations—in terms of the relationships between those operations and the online communities within which they operate. From a high level, we can categorize the first case study as highly *orchestrated*, the second as *cultivated,* and the third as largely *emergent* and *self-sustaining*. However, we can also see each of these dynamics at play, to some extent, across all three studies. Taken together, these case studies provide rich insight into how information operations manifest within online communities and highlight the diffuse intersections between the work of professional information operators and the activities of a distributed, online crowd.

### 5.1 Strategic Information Operations as Collaborative Work

We conceptualize these intersections as a form of collaborative work, using a sociotechnical perspective [26, 60, 61] to highlight the mechanisms of coordination. In some cases we can see evidence of those collaborations in the digital record (e.g. anti-White Helmets activists and agents of information operation campaigns mobilizing through @mentions to spread disinformation) or infer them by triangulating the publically available social media data with other accounts about the inner workings of the information operators (i.e. within the RU-IRA). But in other cases it is less clear that actors are explicitly coordinating their actions, and instead the collaboration consists of convergent behaviors that reflect a kind of intentional shaping or cultivation of an online community.

That shaping is not uni-directional, but has elements of improvisation as the operators work to both reflect and guide the activities of their host community (e.g. RU-IRA agents using #BlackLivesMatter and #BlueLivesMatter hashtags in their accounts and tweets). This dynamic is not new [9, 13, 67, 72]. For example, in Nazi Germany during World War II, orchestrated propaganda campaigns tapped into grassroots organizing efforts to cultivate a network of neighborhood volunteer propagandists called "The Ring" [13]. The Ring filled an important role in Nazi Germany's propaganda system by adjusting national messages to fit local situations and by serving as an interpersonal channel for spreading certain narratives. Bytwerk captures some of the value of this approach by noting how "Germans knew that party members were obligated to say the right things. It was something else when a shopkeeper or

teacher with no obvious party connection made the point" [13]. Ultimately these ordinary citizens were implicated in the propaganda system, but the system itself initially emerged to help them feel empowered by allowing them to "participate in the great events of the day" [30].

Modern information operations represent an extension of these existing logics, although the material affordances of social media can both broaden the role of public participation and make it more visible. Our case studies show how this participation is taking on new shapes by documenting, for instance, the activities of the PJNET group in Study 1 or the anti-war activists and western "journalists" in Study 2. These examples underscore the fact that online information operations do not excusively follow a top-down model of transmitting disinformation through explicitly coordinated actors, but often rely instead on persuading audiences to become "unwitting agents" (in the problematic terms used by intelligence professionals) [9] or "citizen marketers" [67] who take it upon themselves to spread these messages at the grassroots level. That is to say, information operations in online settings involve work that is interdependent in nature—i.e. consisting of independent contributions by different types of formal and informal workers that ultimately rely on each other. Some of these workers are unpaid grassroot intermediaries (e.g. forum moderators, activist grandmothers, and volunteer organizers) who operate like fans to amplify the operation's messages and occasionally take those messages forward in unpredictable ways, while others are paid professionals (e.g. public relations experts and key opinion leaders) trying to support and direct this peer-to-peer activity.

## 5.2 Implications for Design of Platform Policy: The Trouble with "Coordinated Inauthentic Behavior"

This view of information operations as collaborative work has implications for the design of policy. Jackson et al. [41] have argued for a "revival" of CSCW attention to policy, by which they mean both at the macro-level of law and government as well as the meso-level of the internal policies that private companies use to guide their own actions. Here, we echo that argument and highlight some of the ways that a CSCW perspective on strategic information operations can inform some of the current policy challenges that social media platforms face.

As social media companies have moved to address problems of misinformation, disinformation, and other toxic behaviors on their platforms, many have developed policies for defining actors or behaviors that they find problematic and for taking specific actions to address those actors or behaviors (e.g. account or content removal, reduction in visibility, demonetization, etc.). These companies are increasingly communicating with the public in terms of these policies [e.g. 28, 29, 81, 106, 107]. In dealing with disinformation and information operations specifically, the companies have repeatedly expressed reticence to focus on content [29, 107], often stating that they do not want to be the "arbiters of truth" [48, 80]. This position reveals some of the "knot" that Jackson et al. [41] conceptualize between design, practice, and policy—as the platforms attempt to navigate a difficult compromise in relation to existing legal statutes and dynamic social norms around "freedom of speech" and in the midst of increasing political pressure and criticisms of "political bias" [32]. Instead of content, the platforms have elected to focus on authenticity and behavior. For example, Facebook's policy currently stresses a conceptualization of "coordinated inauthentic behavior" [28].

But the cases we have examined—from a CSCW perspective that highlights the collaborative and participatory nature of online information operations—signal some inherent difficulties with such approaches. In particular, our work reveals entanglements between orchestrated action and organic activity, including the proliferation of authentic accounts (real people, sincerely participating) within activities that are guided by and/or integrated into disinformation campaigns. Aligning with research on how active participation strengthens attitudes [62], across our studies, we observe ordinary members of the online crowd appearing, over time, to increasingly view their participation around spreading preferred narratives—sharing anti-Hillary content in the "left" cluster in Study 1, disseminating anti-White Helmets content in Study 2, sharing conspiracy theories of crisis events in Study 3—not only in terms of shaping or influencing outcomes, but as a statement of their own identity. This was made visible in expressive practices like using protest symbols and hashtags in their social media profiles [87], and it is a dynamic that can be seen in other cases of online activism [85]. It is implicated in some of the convergence between how online volunteers and information operators present themselves in their profiles. And, it suggests that though the public is in theory informationally downstream from the messages being propagated by information operations, its "organic" participation

can grow in importance and visibility relative to "orchestrated" efforts. The case of the conspiracy theorizers in Study 3, and to some extent the anti-White Helmets activists in Study 2, speak to this idea of how disinformation can take root in online communities and become, at least partially, self-sustaining. Platform policies designed around rooting out "coordinated inauthentic behavior" would have difficulty addressing these campaigns once they have reached this level of maturity.

Further complicating this picture is the fact that these information operations are themselves shaped by organic activity—as operators routinely "take up" and reflect back messages originating in the crowd. Ong and Cabanes hint at this dynamic in their study of networked disinformation in the Philippines [59], and we see it again in our case study of the disinformation campaign targeting the White Helmets (Study 2). In both cases, activists and operators used overlapping tools and practices to increase outreach—such as targeting specific audiences, harnessing word-of-mouth endorsements for strategic purposes, selectively forwarding news in ways that blur the boundaries between education and persuasion, and engaging not only in 'rational' dialog but also modeling a contagious emotional enthusiasm for others to follow [59, 67]. This perspective again demonstrates the difficulty of differentiating between the orchestrated behaviors of strategic information operators and the organic behaviors of the online communities that those operators routinely target.

These insights suggest that policies built solely on coordination and inauthenticity may fall short of addressing information operations, especially once their work has taken root. They also shed light onto some of the challenges that platforms face in developing and implementing their policies. For example, a more robust approach might consider information operations at the level of a campaign and problematize content based on the *strategic intent of that campaign*. A policy like this could empower the platforms to take action based on the provenance of information—e.g. within a campaign designed to mislead for political purposes or to undermine the integrity of information spaces—rather than the truth value of a piece of content or the authenticity/sincerity of a specific account. However, this approach leaves the platforms in a position of taking action to remove or reduce visibility of content that may be shared or even produced by authentic accounts of sincere online activists. And this might put the platform's policy at odds with commonly held values like "freedom of speech" and platform goals such as providing a place for activists (including those in oppressed groups) to congregate and organize.

These are not simple challenges and they have been shown to be resistent to simple solutions. In addressing them, we argue that platform policies at the intersection of digital expression and strategic information operations cannot be "neutral" or narrowly procedural; they must be substantive, which is to say that they must do their best to ensure that the values that animate our societies are faithfully translated to the digital environment. Crafting such policies will involve challenging conversations that must avoid magical thinking of the sort that frequently enters into technology policy debates (e.g. the fetishization of new technologies or "disruptive innovation"). Echoing Jackson and colleagues' encouragement [41], we argue here that CSCW researchers should be participating in these conversations, helping to guide the development of social media platform policies as they attempt to address these emerging and dynamic threats to the platforms themselves and democratic discourse more broadly.

## 5.3 Implications for Researchers: Theorizing the Effects of Online Information Operations

Currently, there is still substantial uncertainty and speculation regarding the potential impact of information operations involving social media platforms. Unlike the distribution of material artifacts such as pamphlets and posters, the circulation of content on social media creates digital traces that can be systematically analyzed to map the reach of disinformation spreading activities in superb detail. However, the intensely participatory nature of strategic information operations that we have highlighted suggests that it is not enough to see how many people were exposed to any given piece of misleading information; we must understand what these audiences do with it.

Here, we believe it is important to move beyond questions that focus on the highest-level correlations afforded by the analytics of large volumes of data. Such work is welcome and relevant but is only the beginning of a series of inquiries that need to be made into audience interactions with information operations. Computational techniques for analyzing large datasets can allow us to do things like

calculate the viral popularity of deceptive content and correlate it with quantifiable short-term variables such as the number of reshares or election results, but by themselves these approaches are less well equipped to measure complex processes of political identity formation and long-term shifts in cultural and social norms that are part of the logics of techniques like *dezinformatsiya*.

For example, researchers have recently begun analyzing large social media datasets to try to measure the effects of Russian interference in the 2016 U.S. election (which factored in to both our first and third case studies). Some of these recent studies [e.g. 79, 105] have reported observing limited effects on the short-term online behaviors of social media users that were exposed to these information operations. However, care should be exercised when interpreting or framing such findings to avoid suggesting that the effects of information operations are minor, or indeed can even be properly determined. Care is particularly important with the public communications of such findings because of how they can be appropriated by non-experts to inform decision making, as well as political and ethical thinking, in problematic ways [e.g. 51]. While it is indeed a possibility that the effects of information operations are minor, it is also possible that these effects are extremely difficult to measure due to the diffuse nature of strategic information operations, the complex interplay between orchestrated and organic action, and the potential second-order effects that occur through changes to the social networks, social and political norms, and other sociotechnical structures in the broader information space.

This difficulty in quantifying the effects of propaganda is not new. Bittman noted it, while describing the impacts of KGB active measures in a pre-online world, writing: "The results are not statistically measurable, of course, but the KGB evaluates these deception games in broader political terms" [9]. Moreover, social scientists have disputed the idea of information campaigns *directly* influencing the public since at least the 1950s. For instance, Katz and Lazarsfeld's *Personal Influence* [46] argued against understanding the effects of such campaigns using one-step cause-and-effect models, finding that the persuasive effect of any media message was strongly mediated by opinion leaders and one's circle of peers. More recently, Marwick has persuasively argued for the value of employing sociotechnical models to understand the effects of problematically deceptive information, while drawing upon active audience paradigms that reject the idea of treating social media audiences as an undifferentiated mass of interchangeable "cultural dupes" [54].

In principle, CSCW and HCI driven approaches should figure strongly in illuminating some of these dynamics concerning the effects of information operations. CSCW research values working with "Big Data" to understand phenomena that emerge as a consequence of large-scale group activities. At the same time, the field recognizes the importance of understanding "small data", that is the rich, qualitative experiences of social actors who engage in cultural and political practices. Understanding how individuals and groups make meaning from these practices and interpret their political relevance will allow us to develop more robust perspectives on the effects of information operations.

As a closing remark, we offer this reflection, observation and appeal: Not only can the evolving discourse around information operations benefit from CSCW perspectives, CSCW perspectives can benefit from grappling with these murky and muddy phenomena. Information operations both can and have used the pro-social cooperative work taking place over social media as a point of access. This includes, for instance, grassroots political organizing, digital volunteerism, and the wider work of sensemaking and sharing information during breaking news events and crisis situations [e.g. 85, 87, 98, 104]. By considering the dynamics of information operations, we might ourselves be more innovative and understanding of other social phenomena that we are already studying. Our research community is also invested in how we can support processes of being more self-critical and aware as we build and use systems for groups [e.g. 5, 73, 76]. Though information operations are politically charged and thus difficult to study from a "neutral" point of view, they can help us expose and interrogate assumptions and value-commitments hidden to us under normal conditions that nevertheless inform our evaluation and design efforts. Information operations are thus not only important to the study of online collaborations; they are also important contexts from which to study online collaborations.

## 6 CONCLUSION

In this paper, we have framed information operations as a CSCW concern—one with critical societal import. Drawing on case studies and historical context, we show that the work of online information operations extends beyond the narrow window of automated or paid actors such as bots and hired trolls—the window that tends to get the most media and research attention. Our work underscores the

fact that these operations are participatory, taking shape and persisting as collaborations between orchestrated agents and organic crowds. We demonstrate, from a sociotechnical perspective, how information operations function at multiple levels, by directly shaping actions perhaps, but more profoundly by shaping the surrounding social structures—e.g. the networks of activists that take up their messages, along with their norms, practices, and ideologies. We note how these operations take advantage of and resonate with the design of social media platforms that have become central to how information is created, shared, and negotiated across the globe. These platforms face mounting challenges to counter these emerging and dynamic threats, and we argue that a CSCW perspective can help guide them as they develop policies in response. We conclude by highlighting the need for more robust, cross-disciplinary treatments of information operations to understand their emergent features in a technologically mediated world.

## REFERENCES


[1] Norah Abokhodair, Daisy Yoo, and David W. McDonald. 2016. Dissecting a Social Botnet: Growth, Content and Influence in Twitter. In *Proceedings of the 2016 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '16).* ACM, New York, NY, USA, 839-851. DOI: https://doi.org/10.1145/2675133.2675208

[2] Ahmer Arif, John J. Robinson, Stephanie A. Stanek, Elodie S. Fichet, Paul Townsend, Zena Worku, and Kate Starbird. 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 (CSCW '17)*. ACM, New York, NY, 155-168. DOI: https://doi.org/10.1145/2998181.2998294

[3] Ahmer Arif, Leo G. Stewart, and Kate Starbird. (2018). Acting the Part: Examining Information Operations within #BlackLivesMatter Discourse. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW, Article 20 (November 2018), 27 pages. DOI: https://doi.org/10.1145/3274289

[4] Dan Arnaudo. 2017. *Computational Propaganda in Brazil: Social Bots during Elections*. Computational Propaganda Research Project. Oxford University, Oxford, UK.

[5] Karla Badillo-Urquiola, Yaxing Yao, Oshrat Ayalon, Bart Knijnenurg, Xinru Page, Eran Toch, Yang Wang, and Pamela J. Wisniewski. 2018. Privacy in Context: Critically Engaging with Theory to Guide Privacy Research and Design. In *Proceedings of the 2018 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '18).* ACM, New York, NY, 425-431. DOI: https://doi.org/10.1145/3272973.3273012

[6] Ronald M. Baecker, Jonathan Grudin, William A. S. Buxton, and Saul Greenberg (Ed.). 1995. *Groupware and social dynamics: Eight challenges for developers.* Readings in Human–Computer Interaction. DOI: https://doi.org/10.1016/B978-0-08-051574-8.50079-0.

[7] Christine A. Barry, Nicky Britten, Nick Barber, Colin Bradley, and Fiona Stevenson. 1999. Using reflexivity to optimize teamwork in qualitative research. *Qualitative health research* 9, no. 1 (1999): 26-44.

[8] Ladislav Bittman. 1972. *The Deception Game: Czechoslovak Intelligence in Soviet Political Warfare*. Syracuse University Press, Syracuse. NY.

[9] Ladislav Bittman. 1985. *The KGB and Soviet Disinformation: An Insider's View*. Pergamon-Brassey's, Washington, DC.

[10] Lindsay Blackwell, Jill Dimond, Sarita Schoenebeck, and Cliff Lampe. 2017. Classification and Its Consequences for Online Harassment: Design Insights from HeartMob. In *Proceedings of the ACM on Human Computer Interaction 1, CSCW*, Article 24 (November 2017). DOI: https://dx.doi.org/10.1145/3134659

[11] Samantha Bradshaw and Phillip N. Howard. 2018. *Challenging truth and trust: A global inventory of organized social media manipulation.* Computational Propaganda Research Project. Oxford University, Oxford, UK.

[12] Samantha Bradshaw and Phillip N. Howard. 2018. *Why does Junk News Spread So Quickly Across Social Media? Algorithms, Advertising and Exposure in Public Life*. Knight Foundation. Retrieved from https://kf-site-production.s3.amazonaws.com/media_elements/files/000/000/142/original/Topos_KF_ White-Paper_Howard_V1_ado.pdf

[13] Randall L. Bytwerk. 2010. Grassroots propaganda in the Third Reich: The Reich ring for National Socialist propaganda and public enlightenment. *German Studies Review* 33 (2010). 93-118. DOI: http://dx.doi.org/10.2307/40574929

[14] Carole Cadwalladr. 2017. The great British Brexit robbery: how our democracy was hijacked. (May 2017). Retrieved April 3, 2019 from https://www.theguardian.com/technology/2017/may/07/the-great-british-brexit-robbery-hijacked-democracy

[15] Kathy Charmaz. 2014. *Constructing Grounded Theory*. Sage, Thousand Oaks, CA.

[16] Norman K. Denzin and Yvonna S. Lincoln (Eds.). 1994. *Grounded Theory Methodology: An Overview*. Anselm Strauss and Juliet Corbin. Handbook of Qualitative Research. Sage, Thousand Oaks, CA.

[17] Renee DiResta, Dr. Kris Shaffer, Becky Ruppel, David Sullivan, Robert Matney, Ryan Fox, Dr. Jonathan Albright, and Ben Johnson. *The Tactics & Tropes of the Internet Research Agency.* New Knowledge. Retrieved from https://disinformationreport.blob.core.windows.net/disinformation-report/NewKnowledge-Disinformation-Report-Whitepaper.pdf

[18] Paul Dourish and Victoria Bellotti. 1992. Awareness and Coordination in Shared Workspaces. In *Proceedings of the 1992 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '92)*. ACM, New York, NY, USA, 107-114 . DOI: https://doi.org/10.1145/143457.143468

[19] Adam B. Ellick and Adam Westbrook. 2018. Operation Infektion. Russian Disinformation: From Cold War to Kanye. Retrieved April 3, 2019 from https://www.nytimes.com/2018/11/12/opinion/russia-meddling-disinformation-fake-news-elections.html

[20] English Oxford Living Dictionaries. Disinformation. Retrieved April 3, 2019 from https://en.oxforddictionaries.com/definition/disinformation

[21] Robert M. Faris, Hal Roberts, Bruce Etling, Nikki Bourassa, Ethan Zuckerman, and Yochai Benkler. 2017. *Partisanship, Propaganda, and Disinformation: Online Media and the 2016 U.S. Presidential Election*. Berkman Klein Center for Internet & Society Research. Harvard University, Cambridge, Massachusetts, United States.

[22] Emilio Ferrara, Onur Varol, Clayton Davis, Filippo Menczer, and Alessandro Flammini. 2016. The Rise of Social Bots. *Communications of the ACM* 59, 7 (2016). 96-104. DOI: https://doi.org/10.1145/2818717

[23] Vijaya Gadde and Yoel Roth. 2018. Enabling further research of information operations on Twitter. (October 2018). Retrieved April 3, 2019 from https://blog.twitter.com/en_us/topics/company/2018/enabling-further-research-of-information-operations-on-twitter.html

[24] Clifford Geertz. 1975. *The interpretation of cultures*. Chicago: Chicago University Press.

[25] R. Stuart Geiger and David Ribes. 2011. Trace ethnography: Following coordination through documentary practices. In *Proceedings of the Annual Hawaii International Conference on System Sciences*. 1-10. DOI: https://doi.org/10.1109/HICSS.2011.455

[26] Anthony Giddens. 1984. *The Constitution of Society: Outline of the Theory of Structuration*. University of California Press, Berkeley, CA.

[27] Sam Gill. 2019. The price of progress: How "Digital Pollution" is poisoning democracy and what we can do about it. (January 2019). Retrieved April 3, 2019 from https://knightfoundation.org/articles/the-price-of-progress-how-digital-pollution-is-poisoning-democracy-and-what-we-can-do-about-it

[28] Nathaniel Gleicher. 2019. Coordinated Inauthentic Behavior Explained. Facebook Newsroom. (December 2018). Retrieved June 26, 2019 from https://newsroom.fb.com/news/2018/12/inside-feed-coordinated-inauthentic-behavior/

[29] Nathaniel Gleicher. 2019. Removing Coordinated Inauthentic Behavior and Spam From India and Pakistan. (April 2019) Retrieved April 3, 2019 from https://newsroom.fb.com/news/2019/04/cib-and-spam-from-india-pakistan/

[30] Joseph Goebbels. 1938. *Der Rundfunk als achte Großmacht. Signale der neuen Zeit*. 25 ausgewählte Reden von Dr. Joseph Goebbels. 197-207.

[31] Marilys Guillemin and Lynn Gillam. 2004. Ethics, reflexivity, and "ethically important moments" in research. *Qualitative inquiry,* 10.2 (2004): 261-280.

[32] Jessica Guynn. 2019. Ted Cruz threatens to regulate Facebook, Google and Twitter over charges of anti-conservative bias. USA Today. (April 2019). Retrieved June 26, 2019 from https://www.usatoday.com/story/news/2019/04/10/ted-cruz-threatens-regulate-facebook-twitter-over-alleged-bias/3423095002/

[33] Jürgen Habermas. 1984. *The theory of communicative action*, Vol. 2. Beacon press, Boston, MA.

[34] Rongbin Han. 2015. Manufacturing consent in cyberspace: China's 'fifty-cent army'. *Journal of Current Chinese Affairs* 44, 2 (2015). 105-134.

[35] Del Harvey and David Gasca. 2018. Serving Healthy Conversation. Twitter Blog. (May 2018). Retrieved June 26, 2019 from https://blog.twitter.com/en_us/topics/product/2018/Serving_Healthy_Conversation.html

[36] Philip N. Howard. 2002. Network ethnography and the hypermedia organization: New media, new organizations, new methods. *New Media & Society* 4, 4 (2002). 550-574. DOI: https://doi.org/10.1177/146144402321466813

[37] Y. Linlin Huang, Kate Starbird, Mania Orand, Stephanie A. Stanek, and Heather T. Pedersen. 2015. Connected through crisis: Emotional proximity and the spread of misinformation online. In *Proceedings of the 2015 ACM Conference on Computer-Supported Cooperative Work (CSCW '15).* ACM, New York, USA, 969-980. DOI: https://doi.org/10.1145/2675133.2675202

[38] Amanda Lee Hughes and Leysia Palen. 2009. Twitter adoption and use in mass convergence and emergency events. *International Journal of Emergency Management* 6, 3-4 (2009). 248-260.

[39] Edwin Hutchins. 1995. *Cognition in the Wild.* MIT press, Cambridge, MA.

[40] Caroline Jack. 2017. Lexicon of lies: Terms for problematic information. Data & Society. Retrieved from https://datasociety.net/output/lexicon-of-lies/

[41] Steven J. Jackson, Tarleton Gillespie, and Sandy Payette. 2014. The policy knot: Re-integrating policy, practice and design in CSCW studies of social computing. In *Proceedings of the 17th ACM conference on*

*Computer supported cooperative work & social computing (CSCW '14)*, ACM, New York, USA, 588-602. DOI: https://doi.org/10.1145/2531602.2531674
[42] Kathleen Hall Jamieson. 2018. Cyberwar: How Russian Hackers and Trolls Helped Elect a President What We Don't, Can't, and Do Know. Oxford University Press, 2018.
[43] Shagun Jhaver, Sucheta Ghoshal, Amy Bruckman, and Eric Gilbert. 2018. Online harassment and content moderation: The case of blocklists. *ACM Transactions on Computer-Human Interaction (TOCHI 2018)* 25, no. 2 (2018):12. DOI: https://doi.org/10.1145/3185593
[44] Joint Chiefs of Staff. 2014. Information Operations. Joint Publication 3-13. Department of Defense, United States. Retrieved from http://www.jcs.mil/Portals/36/Documents/Doctrine/pubs/jp3_13.pdf
[45] Simon M. Kaplan, William J. Tolone, Douglas P. Bogia, and Celsina Bignoli. 1992. Flexible, Active Support for Collaborative Work with Conversation Builder. In *Proceedings of the 1992 ACM Conference on Computer-Supported Cooperative Work (CSCW '92).* ACM, New York, USA, 378-385. DOI: https://doi.org/10.1145/143457.143552
[46] Elihu Katz, Paul F. Lazarsfeld, and Elmo Roper. 2017. Personal influence: The part played by people in the flow of mass communications. Routledge, London, UK.
[47] Gary King, Jennifer Pan, and Margaret E. Roberts. 2017. How the Chinese government fabricates social media posts for strategic distraction, not engaged argument. *American Political Science Review* 111, 3 (2017). 484-501.
[48] Arjun Kharpal. 2017. Facebook doesn't want to be the 'arbiter of the truth', top exec Sheryl Sandberg says, amid fake news criticism. CNBC.com. (April 2017). Retrieved June 26, 2019 from https://www.cnbc.com/2017/04/24/facebook-fake-news-sheryl-sanberg.html
[49] Audrey M. Kleinsasser. 2000. Researchers, reflexivity, and good data: Writing to unlearn. *Theory into practice* 39.3 (2000): 155-162.
[50] Matthew Levinger. 2018. Master Narratives of Disinformation Campaigns. *Journal of International Affairs* 71, 1.5 (2018). 125-134.
[51] John Leyden. 2018. Kremlin social media trolls aren't actually that influential, study finds. (Jan 2018). Retrieved April 3, 2019 from https://www.theregister.co.uk/2018/01/30/russian_troll_influence/
[52] Herbert S. Lin and Jaclyn Kerr. 2017. *On Cyber-Enabled Information/Influence Warfare and Manipulation*. Oxford University Press, UK.
[53] Jim Maddock, Kate Starbird, Haneen J. Al-Hassani, David E. Sandoval, Mania Orand, and Robert M. Mason. 2015. Characterizing online rumoring behavior using multi-dimensional signatures. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing.* ACM, New York, NY, 228-241. DOI: https://doi.org/10.1145/2675133.2675280
[54] Alice E. Marwick. 2018. Why do People Share Fake News? A Sociotechnical Model of Media Effects. *Georgetown Law Technology Review* 2, 474 (2018). 474-512.
[55] Alice E. Marwick and danah boyd. 2011. I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New Media & Society* 13, 1 (2011), 114-133. DOI: https://doi.org/10.1177/1461444810365313
[56] Jolie Myers and Monika Evstatieva. 2018. Meet The Activist Who Uncovered The Russian Troll Factory Named In The Mueller Probe. NPR: Parallels—Many Stories, One World. (March 2018). Retrieved June 26, 2016 from https://www.npr.org/sections/parallels/2018/03/15/594062887/some-russians-see-u-s-investigation-into-russian-election-meddling-as-a-soap-ope
[57] Yiftach Nagar. 2012. What do you think?: The Structuring of an Online Community as a Collective-Sensemaking Process. In *Proceedings of the 2012 ACM Conference on Computer-Supported Cooperative Work (CSCW '12).* ACM, New York, USA, 393-402. DOI: https://doi.org/10.1145/2145204.2145266
[58] Office of the Director of National Intelligence. 2017. Assessing Russian activities and intentions in recent US elections. National Intelligence Council. (January 2017). Retrieved April 3, 2019 from https://www.dni.gov/files/documents/ICA_2017_01.pdf
[59] Jonathan Corpus Ong and Jason Vincent A. Cabanes. 2018. Architects of Networked Disinformation. The Newton Tech4Dev Network. University of Leeds, Leeds, UK. Retrieved from http://newtontechfordev.com/wp-content/uploads/2018/02/ARCHITECTS-OF-NETWORKED-DISINFORMATION-FULL-REPORT.pdf
[60] Wanda J. Orlikowski.1992. The duality of technology: Rethinking the concept of technology in organizations. *Organization science* 3, 3 (1992). 398-427.
[61] Wanda J. Orlikowski. 2007. Sociomaterial practices: Exploring technology at work. *Organization studies* 28, 9 (2007). 1435-1448.
[62] Stuart Oskamp and P. Wesley Schultz. 2005. *Attitudes and Opinions* (3rd ed.). Lawrence Erlbaum, Mahwah, NJ.
[63] Leysia Palen and Kenneth M. Anderson. 2016. Crisis informatics - New data for extraordinary times. *Science* 353, 6296 (2016). 224-225. DOI: https://doi.org/10.1126/science.aag2579

[64] Leysia Palen and Sophia B. Liu. 2007. Citizen Communications in Crisis: Anticipating a Future of ICT-Supported Public Participation. In *Proceedings of the 2007 CHI Conference on Human Factors in Computing Systems (CHI '07)*. ACM, New York, NY, 727-736.
[65] Alicia Parlapiano and Jasmine C. Lee. 2018. The propaganda tools used by Russians to influence the 2016 election. (February 2018). Retrieved April 3, 2019 from https://www.nytimes.com/interactive/2018/02/16/us/politics/russia-propaganda-election-2016.html
[66] Christopher Paul and Miriam Matthews. 2016. The Russian "Firehose of Falsehood" Propaganda Model. Rand Corporation, Santa Monica, CA. DOI: https://doi.org/10.7249/PE198
[67] Joel Penney. 2017. The citizen marketer: Promoting political opinion in the social media age. Oxford University Press, Oxford, UK.
[68] Peter Pomerantsev. The Kremlin's information war. *Journal of Democracy* 26, 4 (2015). 40-50.
[69] Peter Pomerantsev and Michael Weiss. 2014. *The menace of unreality: How the Kremlin weaponizes information, culture and money.* Institute of Modern Russia, New York, NY.
[70] David Rees. 1984. *Soviet Active Measure: The Propaganda War*. Institute for the Study of Conflict, London, UK.
[71] Aleksandra Sarcevic, Leysia Palen, Joanne White, Kate Starbird, Mossaab Bagdouri, and Kenneth Anderson. 2012. "Beacons of Hope" in Decentralized Coordination: Learning from On-the-Ground Medical Twitterers During the 2010 Haiti Earthquake. In *Proceedings of the 2012 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '12).* ACM, New York, NY, 47-56. DOI: https://doi.org/10.1145/2145204.2145217
[72] Jeffrey K. Sawyer. 1990. Printed poison: Pamphlet propaganda, faction politics, and the public sphere in early seventeenth-century France. Univ of California Press, Berkeley, CA.
[73] Phoebe Sengers, John McCarthy, and Paul Dourish. 2006. Reflective HCI: articulating an agenda for critical practice. In *Proceedings of the CHI'06 extended abstracts on Human Factors in Computing Systems*. ACM, New York, NY, 1683-1686. DOI: https://doi.org/10.1145/1125451.1125762
[74] David K. Shipler. 1986. After they defect... (December 1986). Retrieved April 3, 2019 from https://www.nytimes.com/1986/12/07/magazine/after-they-defect.html
[75] Craig Silverman. 2018. Russian Trolls Ran Wild On Tumblr And The Company Refuses To Say Anything About It. (Feb. 2018). Retrieved April 17, 2018 from https://www.buzzfeed.com/craigsilverman/russian-trolls-ran-wild-on-tumblr-and-the-company-refuses?utm_term=.ad65gb5jz#.rdwOw8O6Z
[76] Robert Soden and Leysia Palen. 2018. Informating Crisis: Expanding Critical Perspectives in Crisis Informatics. In *Proceedings of the ACM on Human-Computer Interaction 2,* CSCW, Article 162 (November 2018). DOI: https://doi.org/10.1145/3274431
[77] Olivia Solon. 2017. How Syria's White Helmets became victims of an online propaganda machine. (December 2017). Retrieved April 3, 2019 from https://www.theguardian.com/world/2017/dec/18/syria-white-helmets-conspiracy-theories
[78] Andrei Soshnikov. 2015. The capital of political trolling. MR7.ru. (March 2015). Retrieved June 26, 2016 from https://mr-7.ru/articles/112478/
[79] Alexander Spangher, Gireeja Ranade, Besmira Nushi, Adam Fourney, and Eric Horvitz. 2018. Analysis of Strategy and Spread of Russia-sponsored Content in the US in 2017. arXiv:1810.10033. Retrieved from https://arxiv.org/abs/1810.10033
[80] Supraja Srinivasan. 2018. We don't want to be arbiters of truth: YouTube CBO Robert Kyncl. Economic Times/India Times. (March 2018). Retrieved June 26, 2019 from https://tech.economictimes.indiatimes.com/news/internet/we-dont-want-to-be-arbiters-of-truth-youtube-cbo-robert-kyncl/63438805
[81] Alex Stamos. 2018. Authenticity Matters: The IRA Has No Place on Facebook. (Apr. 2018) Retrieved April 17, 2018 from https://newsroom.fb.com/news/2018/04/authenticity-matters/
[82] Kate Starbird. 2013. Delivering Patients to Sacré Coeur: Collective Intelligence in Digital Volunteer Communities. In *Proceedings of the 2013 CHI Conference on Human Factors in Computing Systems (CHI '13).* ACM, New York, NY, 801-810. DOI: https://doi.org/10.1145/2470654.2470769
[83] Kate Starbird. 2017. Examining the Alternative Media Ecosystem through the Production of Alternative Narratives of Mass Shooting Events on Twitter. In *11th International AAAI Conference on Web and Social Media (ICWSM 2017)*, Montreal, Canada, (pp. 230-339).
[84] Kate Starbird, Ahmer Arif, Tom Wilson, Katherine Van Koevering, Katya Yefimova, and Daniel Scarnecchia. (2018). Ecosystem or Echo-System? Exploring Content Sharing across Alternative Media Domains. In *12th International AAAI Conference on Web and Social Media (ICWSM 2018)*, Stanford, CA, (pp. 365-374).
[85] Kate Starbird and Leysia Palen. 2011. Voluntweeters: Self-Organizing by Digital Volunteers in Times of Crisis. In *Proceedings of the 2011 CHI Conference on Human Factors in Computing Systems (CHI '11)*. ACM, New York, NY, 1071-1080. DOI: https://doi.org/10.1145/1978942.1979102
[86] Kate Starbird and Leysia Palen. 2013. Working and Sustaining the Virtual Disaster Desk. In *Proceedings of the 2013 ACM Conference on Computer-Supported Cooperative Work (CSCW '13).* ACM, New York, USA, 491-502. DOI: https://doi.org/10.1145/2441776.2441832
[87] Leo Graiden Stewart, Ahmer Arif, A. Conrad Nied, Emma S. Spiro, and Kate Starbird. 2017. Drawing the lines of contention: Networked frame contests within #BlackLivesMatter discourse. In *Proceedings of ACM*

*Human-Computer Interaction 1*, CSCW, Article 96 (December 2017), 23 pages. DOI: https://doi.org/10.1145/3134920
[88] Cass R. Sunstein. 2014. *Conspiracy Theories and Other Dangerous Ideas*. Simon and Schuster, New York, NY.
[89] Jonathan Supovitz, Alan J. Daly, Miguel del Fresno, and Christian Kolouch. 2017. #CommonCore Project. Retrieved April 3, 2019 from www.hashtagcommoncore.com
[90] The Internet Archive. About the Internet Archive. Retrieved April 17, 2018 from https://archive.org/about/
[91] Rod Thornton. 2007. Asymmetric warfare: Threat and response in the 21st century. Polity, Cambridge, UK.
[92] Tumblr Help Center. 2018. Public Record of Usernames Linked to State-Sponsored Disinformation Campaigns. (Mar. 2018). Retrieved April 17, 2018 from https://tumblr.zendesk.com/hc/en-us/articles/360002280214
[93] Twitter. 2018. Update on Twitter's Review of the 2016 U.S. Election. (Jan. 2018) Retrieved April 17, 2018 from https://blog.twitter.com/official/en_us/topics/company/2018/2016-election-update.html
[94] United States House of Representatives Permanent Select Committee on Intelligence. 2017. Exhibit B (Nov. 2017). https://democrats-intelligence.house.gov/uploadedfiles/exhibit_b.pdf
[95] Sarah Vieweg, Amanda L. Hughes, Kate Starbird, and Leysia Palen. 2010. Microblogging During Two Natural Hazards Events: What Twitter May Contribute to Situational Awareness. In *Proceedings of the 2010 CHI Conference on Human Factors in Computing Systems (CHI '10)*. ACM, New York, NY, 1079-1088. DOI: https://doi.org/10.1145/1753326.1753486
[96] Morgan Vigil-Hayes, Marisa Duarte, Nicholet Deschine Parkhurst, and Elizabeth Belding. 2017. #Indigenous: Tracking the Connective Actions of Native American Advocates on Twitter. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '17).* ACM, New York, NY, USA, 1387-1399. DOI: https://doi.org/10.1145/2998181.2998194
[97] Jessica Vitak, Kalyani Chadha, Linda Steiner, and Zahra Ashktorab. 2017. Identifying Women's Experiences With and Strategies for Mitigating Negative Effects of Online Harassment. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '17)*. ACM, New York, USA, 1231-1245. DOI: https://doi.org/10.1145/2998181.2998337.
[98] Shaun Walker. 2017. Russian troll factory paid US activists to help fund protests during election. The Guardian. Available at: https://www.theguardian.com/world/2017/oct/17/russian-troll-factory-activists-protests-us-election
[99] Yiran Wang and Gloria Mark. 2017. Engaging with Political and Social Issues on Facebook in College Life. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '17).* ACM, New York, NY, USA, 433-445. DOI: https://doi.org/10.1145/2998181.2998295
[100] Clair Wardle and Hossein Derakhshan. 2017. Information Disorder: Toward an Interdisciplinary Framework for Research and Policy Making. Council of Europe Report.
[101] Jen Weedon, William Nuland and Alex Stamos. 2017. Information Operations and Facebook. (Apr. 2017) Retrieved March 17th, 2019 from https://fbnewsroomus.files.wordpress.com/2017/04/facebook-and-information-operations-v1.pdf
[102] Tom Wilson, Kaitlyn Zhou, and Kate Starbird. 2018. Assembling Strategic Narratives: Information Operations as Collaborative Work within an Online Community. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW, Article 183 (November 2018), 26 pages. DOI: https://doi.org/10.1145/3274452
[103] Samuel C. Woolley and Philip N. Howard. 2016. Political communication, computational propaganda, and autonomous agents: Introduction. *International Journal of Communication* 10 (2016).
[104] Volker Wulf, Kaoru Misaki, Meryem Atam, and David Randall. 2013. 'On the ground'in Sidi Bouzid: Investigating Social Media Use During the Tunisian Revolution. In *Proceedings of the 2013 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '13).* ACM, New York, USA, 1409-1418. DOI: https://doi.org/10.1145/2441776.2441935
[105] Savvas Zannettou, Tristan Caulfield, Emiliano De Cristofaro, Michael Sirivianos, Gianluca Stringhini, and Jeremy Blackburn. 2018. Disinformation Warfare: Understanding State-Sponsored Trolls on Twitter and Their Influence on the Web. arXiv:1801.09288. Retrieved from https://arxiv.org/abs/1801.09288
[106] YouTube. 2019. Our Ongoing Work to Tackle Hate. Official Blog. (June 2019). Retrieved June 26, 2019 from https://youtube.googleblog.com/2019/06/our-ongoing-work-to-tackle-hate.html
[107] Mark Zuckerberg. 2019. Mark Zuckerberg: The Internet needs new rules. Let's start in these four areas. (March 2019). Retrieved April 3, 2019 from https://www.washingtonpost.com/opinions/mark-zuckerberg-the-internet-needs-new-rules-lets-start-in-these-four-areas/2019/03/29/9e6f0504-521a-11e9-a3f7-78b7525a8d5f_story.html