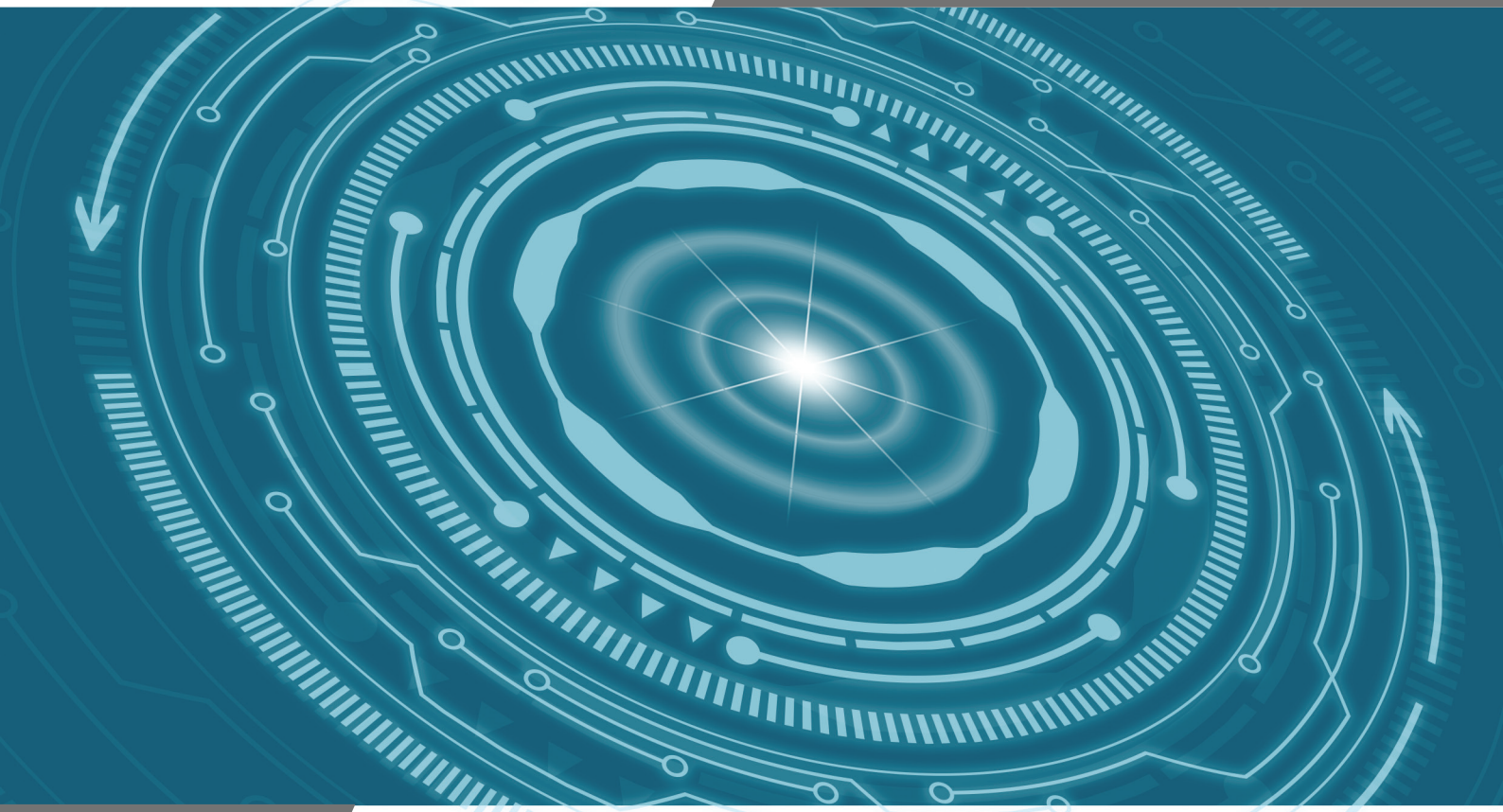
Centre for Economics and Foreign Policy Studies

edam

Cyber Governance and Digital Democracy 2019/3



November 2019

HOSTILE INFLUENCE AND EMERGING COGNITIVE THREATS IN CYBERSPACE

Baris Kirdemir | EDAM & R. Bosch Cyber Policy Fellow

HOSTILE INFLUENCE AND EMERGING COGNITIVE THREATS IN CYBERSPACE

Baris Kirdemir | EDAM & R. Bosch Cyber Policy Fellow

INTRODUCTION

Disinformation draws unprecedented public attention. Current digitalization and technological transformation alter how people consume information, perceive the world, and make decisions. A diverse set of actors, ranging from foreign governments to terrorist outlets and fraudsters, use cyber-mediated information operations for a variety of purposes, including gaining political or economic influence. Disinformation and social manipulation through cyber-mediated channels alter basic social mechanisms and threaten foundational democratic structures. Political polarization, radicalization, and violent extremism are now partly connected to informational dynamics across the cyber-space. Authoritarian governments combine new technologies and the features of the new information environment to suppress political opposition, freedom of expression, or certain racial or ethnic groups.

In light of the recent trends and growing knowledge across the scientific and policy research literature, this paper presents an overview of the emerging cyber-mediated security threats as well as their underlying social and cognitive dynamics. The first section delivers an assessment of how the modern information environment and social-political dynamics change vis-à-vis the threat of malign influence across the cyber-space. The second section offers some selected insights from the events and activities on social media platforms, with a specific focus on the factors that reach beyond the basic concept of fake-news. Finally, the third section explores a tiny fraction of the scientific literature to illustrate both social and cognitive features that relate to current and future challenges.

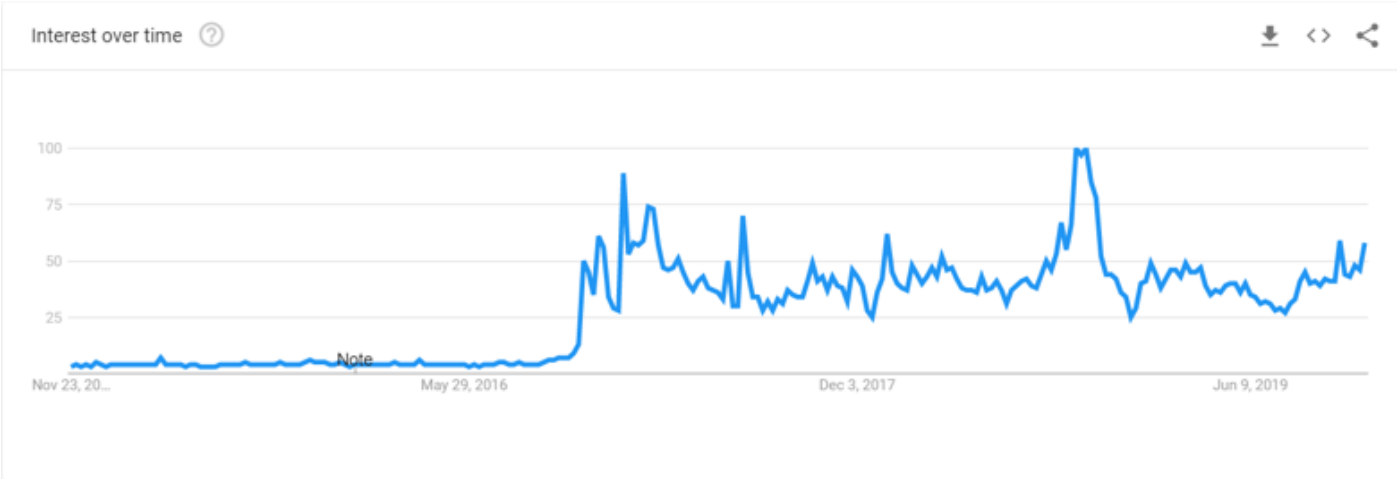
Robert Bosch Stiftung The research has been made possible by funding obtained from Robert Bosch Stiftung.

Across the Cyberspace: Understanding the Threat of Influence

Modern *"information disorder"*[1] causes multiple challenges that extend beyond the so-called problem of "fake news." Cyber-mediated hostile information campaigns aim to alter behavior, attitudes, and perception of reality in targeted societies. They aim to cause or amplify confusion, social unrest, polarization, hatred, violent extremism, and erosion of trust. Beyond common tools such as the fakeness or distortion of facts, manipulative campaigns benefit from how people cope with their exposure to an extreme amount of information on a daily basis.

Regularly bombarded with misleading headlines, statistics, frames, and narratives, human cognition relies on *"mental shortcuts"* to overcome its limitations. Spreaders of disinformation, knowingly or not, often utilize this tendency by imitating legitimacy, impersonating known credible sources, or by using misleading facts and statistics. Besides, the use of emotive content and "cognitive hacks" may alter how people receive the given information. There is a significant level of agreement that, as a result of financial incentives as well as continuous look for online user traffic and engagement, modern news dissemination and consumption habits only add to the problem.[2]

News outlets, social media companies, and non-governmental organizations frequently report new revelations on foreign influence operations. To illustrate, researchers from Princeton University studied publicly reported "foreign influence efforts" to document recent trends.[3] The study identified 53 different foreign influence efforts in 24 targeted countries from 2013 to 2018. Within the set of reported incidents, most common actors that served foreign influence were private companies, media organizations, foreign government officials, and intelligence agencies. Defamation -attacking the reputation of and trust in institutions and people-, as well as persuasion and polarization, were among common strategies, while tactics such as the creation of original misleading content, amplification of existing materials, hijacking conversations, and distortion of the facts evolved in terms of their proportions in time. Attackers used bots and trolls on multiple platforms such as Twitter, Facebook, and news outlets.[4] Beyond the mentioned study, it is probably safe to assume a higher number of foreign influence campaigns at any given time. As authors also suggest, publicly available reports on such campaigns rely on what has been already discovered. Moreover, the operations of certain countries attract more media attention.



"Fake news" topic on Google Trends. "Numbers represent search interest relative to the highest point on the chart for the given region and time. A value of 100 is the peak popularity for the term. A value of 50 means that the term is half as popular. A score of 0 means there was not enough data for this term." Source: Google

1 First Draft, Understanding Information Disorder, 2019.

2 Ibid.

3 Diego A. Martin and Jacob N. Shapiro, Trends in Online Foreign Influence Efforts, ESOC Publications, 2019.

4 For detailed documentation; see; Ibid.

Disinformation is *"the deliberate promotion of false, misleading, or misattributed information in either content or context."*[5] In recent years, tackling the intertwined phenomena of disinformation and social manipulation has become one of the most pressing security policy issues. Both states and non-state actors use new technologies, cyber-mediated communication platforms, and the new social structure to conduct hostile influence operations. However, given the systemic transformation across the modern information environment, hostile manipulative campaigns are likely to grow to much greater levels. Therefore, the problem of social-cognitive manipulation is beyond disinformation and fake news, as it occurs in a rapidly transforming information environment and directly threatens the central pillars of modern societies and political systems.

A social manipulation campaign deploys various tools, usually within a larger political-military context, to influence the perception of reality of the targeted audience. Evaluation of cognitive influence is within the framework of strategic effectiveness and performance.[6] From a political-military perspective, the socio-technological transformation may facilitate real *"information blitzkriegs,"* with potential overarching implications for the global geopolitics.[7] Operating in the information and cognitive dimensions to achieve political objectives with no or minimal use of physical force has become a common characteristic of international conflicts. Moreover, every action or inaction in both physical and cognitive domains are evaluated with regards to their informational utility.

Consequently, warfare is increasingly waged by networks against each other, and the center of gravity for each battle is shifting towards the human mind and cognitive processes. Referring to this transformation, a recent study by the RAND Corporation emphasizes the risk of "virtual societal warfare" as a new form of conflict. The study defines hostile social manipulation as *"the purposeful, systematic generation and dissemination of information to produce harmful social, political, and economic outcomes in a target country by affecting beliefs, attitudes, and behavior."*[8]

Influencing a broad audience is more difficult than smaller groups that maintain certain similarities, such as shared educational or professional background, ideology, and other characteristics. Web-based channels, popular social media platforms, and mobile device applications shift the information dissemination from a broad and relatively centralized context to target audience-specific and decentralized forms. Microtargeting in the commercial sector, personalized content, and recommendation algorithms are among the well-known examples of this shift. Thus, identifying and targeting the interests and personality features of influencers, spreaders, and target audiences are common practices not only in legitimate advertisement campaigns but also in hostile information operations.[9]

Micromarketing and psycho-profiling techniques enable an increasing amount of actors to target specific groups or even individuals with tailored information and "cognitive cues." Psychology, behavioral economics, and machine learning are among the fields that will enable new tools to influence and continuously manipulate people.[10]

Due to the new characteristics of communication, such challenges relate to the cyber-space. The cybersecurity efforts are often related to the physical information networks, software, safety of data from cyber-attacks and stealing, safety of physical infrastructure from damage, information security, and other critical security priorities. Overall, this scope does not cover the human dimension, namely the cognitive, emotional, social, and behavioral aspects that are now integrated into the modern cyber-space.[11]

---

5 Michael Krigsman, Pablo Breuer, Sara-Jayne Terp and David A. Bray, Disinformation, Cognitive Security, and Infleunce, CXOTALK, https://www.cxotalk.com/episode/disinformation-cognitive-security-influence, Accessed on: November 15, 2019.

6 Yossi Kuperwasser and David Siman-Tov (ed), The Cognitive Campaign: Strategic and Intelligence Perspectives, The Institute for National Security Studies and The Institute for the Research of the Methodology of Intelligence, 2019.

7 D.M. Beskow and K.M. Carley Social cybersecurity: an emerging national security requirement. Military Review, 99(2), 117, 2019.

8 Michael J. Mazarr, Ryan Michael Bauer, Abigail Casey, Sarah Heintz and Luke J. Matthews, The Emerging Risk of Virtual Societal Warfare, 2019.

9 Haim Assa, "Influencing Public Opinion", in Yossi Kuperwasser and David Siman-Tov (ed), The Cognitive Campaign: Strategic and Intelligence Perspectives, The Institute for National Security Studies and The Institute for the Research of the Methodology of Intelligence, 2019 , 25-35.

10 Ibid.

11 "Integrating Social and Behavioral Sciences (SBS) Research to Enhance Security in Cyberspace", in National Academies of Sciences, Engineering, and Medicine, A Decadal Survey of the Social and Behavioral Sciences: A Research Agenda for Advancing Intelligence Analysis. Washington, DC: The National Academies Press, 2019.

For example, from the cybersecurity perspective, *"cognitive hacking"* is one of the terms associated with deception and behavioral manipulation, referring to the weaponization of information to induce behavioral changes in targeted humans.[12] Cybersecurity vulnerabilities can be exploited for social manipulation purposes. For instance, the rapid dissemination of misleading content from a hacked social media account of a company may lead to stock price changes and financial losses by imposing behavioral changes on humans who receive the message. In cognitive hacking, *"humans become the tools of attackers."* A combination of tools enabled by machine learning and natural language processing would enable novel attacks. In this context, cognitive hacking would have overarching political, social, and economic implications.[13]

The concept of "netwars," developed in the 1990s simultaneously with the projections of future cyberwars, includes the efforts to effectively shape what a target population knows, perceives, and believes about its environment. Netwars are more relevant to human cognition and emotions, while cyberwar relates to physical networks, infrastructure, military systems, and information security.[14]

Information processing constitutes the very core of society and political systems. Social manipulation may increasingly target this core foundation. People and "machines" become integrated, though in a very fragmentized fashion, in a system mediated by hyper-connectivity and sophisticated algorithms across the cyber-space. This systemic transformation is much faster than humans' evolutionary adaptation, naturally, and understanding how inherent foundations of human cognition would react to the new social-informational structures will be the key to ensure a reasonable level of security in the near future.

Within the given context, people are increasingly exposed to a mixture of decontextualized facts, distortion, deceptive use of statistics, dismissal of issues and narratives, partisan and emotional content, distraction, ethnic or racial bias, and a variety of computational techniques to amplify the effect of such content. Thus, tackling these offenses require efforts beyond fact-checking and exposure of debunked fake news. Overall, these *"information maneuvers"* are increasingly diverse in terms of the set of tactics, techniques, and procedures they employ.[15]

One particular scientific field, already enhancing the knowledge about social, behavioral, and technological dynamics that relate to the overarching problem is computational social science. Computational social science is *"the use of social science theories to drive the development of new computational techniques, combined with further development of those theories using computational techniques for data collection, analysis, and simulation."*[16] Computational social science integrates computer science with social and behavioral sciences, forming a real transdisciplinary field that can fill the knowledge gaps and serve to tackle contemporary policy issues.

More specifically, a younger but rapidly developing research field, being formed to address the cyber-mediated social manipulation is "**social cybersecurity.**" Social cybersecurity is characterized as a sub-field of computational social science. In essence, social cybersecurity is an operational field, and it is connected to various other terms such as *"social cyber forensics, social cyber-attack, social media analytics, computational propaganda, and social media information."*[17] Although it is connected to traditional cybersecurity approaches, social cybersecurity primarily focuses on the human aspect.

The social cybersecurity field develops a wider and deeper understanding of how the abovementioned cyber-mediated challenges impact human beliefs, attitudes, cognition, emotions, and behavior. A significant amount of research effort helps to track, monitor, and understand how social manipulation works. Often, researchers and practitioners from the social cybersecurity field also offer new tools to classify and predict potential threats such as the use of

12 Groh, Selena, Cognitive Hacking How to Fight Fake News, Tufts University, 2017.

13 Ibid.

14 Linton Wells II, Cognitive-Emotional Conflict, PRISM, 7(2), 5, 2017.

15 D.M. Beskow and K.M. Carley Social cybersecurity: an emerging national security requirement. Military Review, 99(2), 117, 2019.

16 "Integrating Social and Behavioral Sciences (SBS) Research to Enhance Security in Cyberspace", in National Academies of Sciences, Engineering, and Medicine, A Decadal Survey of the Social and Behavioral Sciences: A Research Agenda for Advancing Intelligence Analysis. Washington, DC: The National Academies Press, 2019.

17 Ibid.

botnets or other computationally coordinated campaigns. Most importantly, social cybersecurity field enables a conceptual framework to develop and operationalize a large amount of interdisciplinary knowledge from a multitude of older fields such as computer science, information science, psychology, neuroscience, anthropology, biology, social science, political science, and many others in a potentially policy-relevant framework. All in all, social cybersecurity would become an important component of the fusion between the research, industry, civil society, and policy environment, as such an integrative framework is now much needed.

The modern information environment is increasingly defined by the decentralized networks of information sharing and decision-making, combined with the fragmentation of belief structures, alternative realities, and continuous aggression by inauthentic amplifiers such as trolls and bots. The proliferation of new technologies such as artificial intelligence, virtual and augmented reality, the Internet of Things, and others require governments and societies to adapt to this new *"infosphere"* by developing new norms and regulatory frameworks.[18] The adaptability of overall national security structures will be strongly relevant in this context, with profound implications for future conflicts.

Hostile Influence, Radicalization, and Violent Extremism: Insights from Social Media

Social media is an arena of competition, with rivals continuously trying to outpace each other for *"power and influence."*[19] Actors that produce disinformation include *"trolls, bots, fake-news websites, conspiracy theorists, politicians, highly partisan media outlets, the mainstream media, and foreign governments."*[20] Online disinformation campaigns, using trolls, bots, and cyborgs (accounts operated by humans and bots together) as force multipliers, often promote selected information sources above others, increasing their visibility and perceived popularity. Tactics include selective censorship, manipulating search algorithms (mutual admiration societies, keyword stuffing, link bombs, algorithmic manipulation, hijacking hashtags, and conversation), hacking sensitive and damaging information, directly introducing and spreading disinformation.[21] The "incivility" and toxicity of online political information are on the rise. Moreover, studies suggest that the sources of such uncivil content, as well as subsequent comments with high incivility scores, are prone to be more popular among the viewers.[22]

Furthermore, the assessment of previous disinformation campaigns indicates the possibility of specific targeting strategies for different population groups. This is a result of group-specific, nuanced reactions to disinformation. For example, a study on the dissemination of fake news on Facebook during the 2016 Presidential election in the United States found that age is one of the predictors of sharing false information, adding to ideology. People older than 65 were seven times more likely to disseminate fake news, the study showed.[23] Other well-known examples of specifically targeted population groups include highly-partisan groups or certain ethnicities.

Information operations and manipulative campaigns on social media resemble money laundering practices. Networks of fake personas, groups, channels, pages, and websites disseminate disinformation while hiding their relationship, ownership structure, and overall intent. These networks disseminate falsehoods by "laundering" information . Similar to illicit financing, sources of information imitate legitimacy and authenticity. Resembling money laundering circles

18 Michael J. Mazarr, Ryan Michael Bauer, Abigail Casey, Sarah Heintz and Luke J. Matthews, The Emerging Risk of Virtual Societal Warfare, 2019.

19 Joshua A. Tucker, Andrew Guess, Pablo Barberá, Cristian Vaccari, Alexandra Siegel, Sergey Sanovich, Denis Stukal and Brendan Nyhan. Social media, Political Polarization, and Political Disinformation: A Review of the Scientific Literature, 2018.

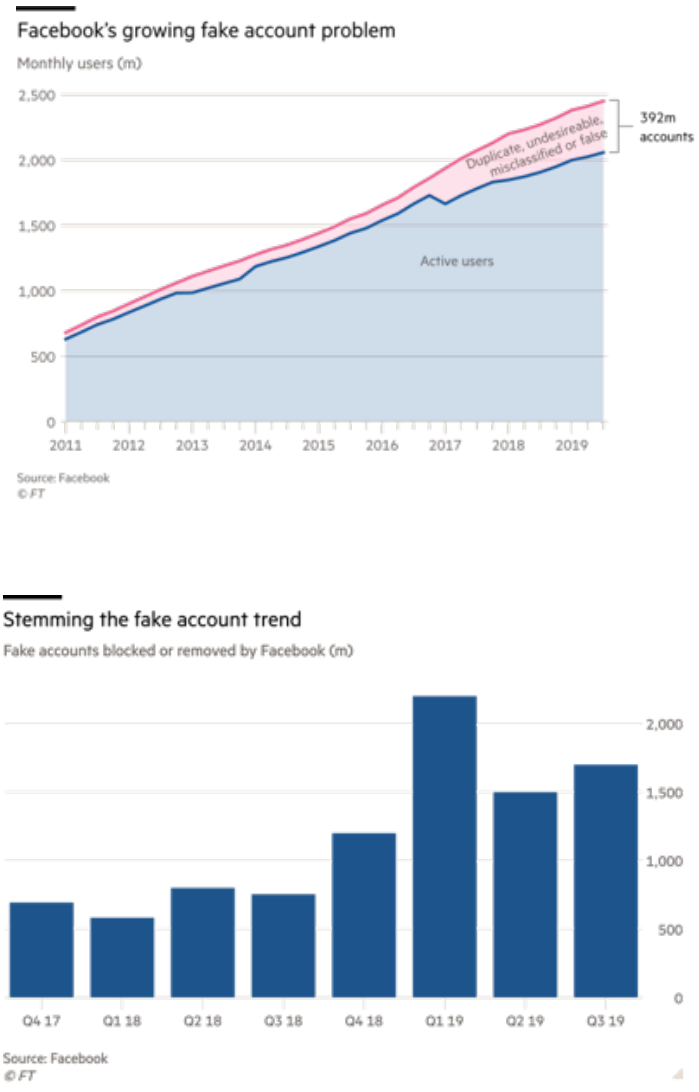
20 Ibid.

21 Ibid.

22 Ibid.

23 A. Guess, J. Nagler, & J. Tucker, Less Than You Think: Prevalence and Predictors of Fake News Dissemination on Facebook. Science Advances, 5(1), 2019.

that split money transfers into many small pieces as fake financial transactions, information operations on social media refer to multiple sources, and the information is disseminated through many different channels while hiding the relationship between these sources and amplifiers.[24] The "black markets" of fake social media accounts and user engagement often facilitate these campaigns.[25] Fake accounts, views, likes, shares, and other user engagement are marketed through a considerably large black market that is easily available to anyone who can use a search engine. The black market services range from providing simple accounts to disinformation pieces in major news outlets. The cost of a purchase ranges from as cheap as a few to thousands of US Dollars.



Number of fake accounts on Facebooks has been growing in recent years. Facebook has removed billions of fake accounts in 2019. Sources: Financial Times[26], Facebook.

Social bots are automated computer programs that create and disseminate content by interacting with other bots or humans on various platforms such as social media or online multiplayer games. Primitive versions of social bots emerged in the 1990s by engaging in simple conversations with people on chat channels. Modern social bots, however, are much more capable than those early versions. Therefore, many legitimate and "benign" variances of social bots facilitate applications by commercial actors, government agencies, or civil society. On the other hand, malicious social bots can pollute the modern information environment in various ways, such as forming networks or infiltrating the existing ones and creating a false perception of the support an actor, idea, or narrative has. Social bots create a false perception in which a certain piece of information or a narrative seems to be coming from many distinct sources. Combined with high volumes of repetitiveness, social bots may increase the probability of alterations in human behavior, beliefs, or attitudes. Moreover, by polluting metrics on social media, social bots challenge large-scale algorithms that constantly check trends and sentiments online.[27]

Social bots have been used for a variety of purposes. Social bot-enabled "Twitter bombs" are used to demobilize and suppress opposing political groups, or to create false perceptions of popular issues, messages, and actors.[28] Both state and non-state actors use social bots to pursue their political agenda. Botnet campaigns are increasingly active during international crises, conflict events, and wars. Networks of social bots are operated in many different ways.

24 Kirill Meleshevich and Bret Schafer, Online Information Laundering: The Role of Social Media, GMF, 2018.

25 NATO Strategic Communications Centre of Excellence, The Black Market for Social Media Manipulation, 2018.

26 Financial Times, https://www.ft.com/content/98454222-fef1-11e9-b7bc-f3fa4e77dd47, Accessed on: 20 November 2019.

27 Emilio Ferrera, Onur Varol, Clayton Davis, Filippo Menczer and Alessandro Flammini, The Rise of Social Bots, Communications of the ACM 59 (7), 96-104, 2016.

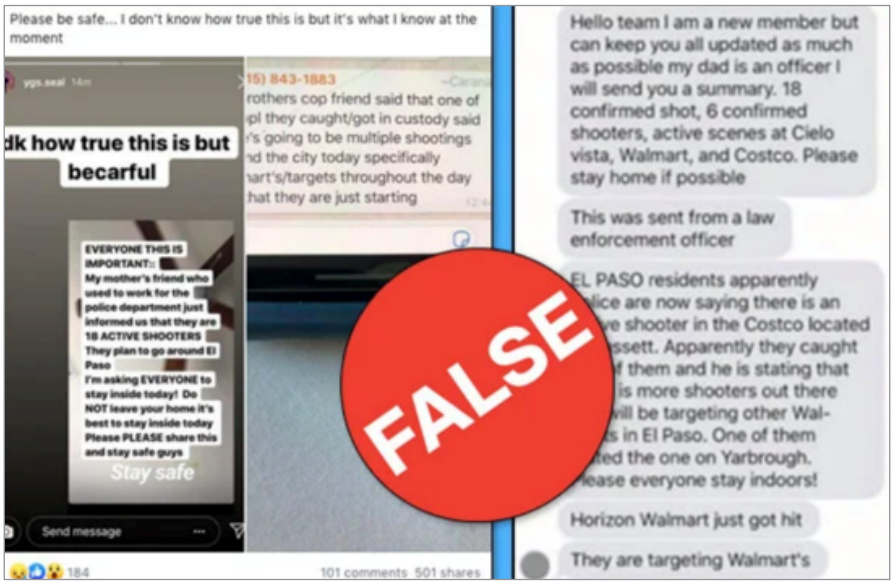
28 Woolley, Samuel C. "Automating Power: Social Bot Interference in Global Politics, First Monday, 21 (4), 2016.

Throughout the conflict in Ukraine, for example, botnets employed a variety of deceptive tactics. These tactics were continuously fine-tuned to influence different targeted groups.[29]

The tech industry and research community constantly develop and upgrade systems to detect bots. Using a variety of tools from fields like machine learning and social network analysis, bot detection systems are proven to be effective in some cases. However, no detection system is perfect, and social bots' strategies continue to evolve. Some botnets are operated by complex networks of humans and automated bots at the same time, and some social bots are more capable of imitating human behavior. The competition between malicious bots and bot detection systems is likely to resemble some arms race and continue to evolve in the foreseeable future.

Disinformation and extremist communication activities have been moving into closed platforms that are harder to monitor. For the research community and non-governmental organizations that track online disinformation, it is significantly more difficult to take a full picture of how, when, and by whom an information operation initiated and evolved once it moved to closed mediums such as WhatsApp, Telegram, or Facebook Messenger.[30] On the other hand, the reach and impact of disinformation probably become more limited, at least until the new platform attracts enough people. Most importantly, the migration of disinformation from major open platforms to closed ones is not absolute. Disinformation usually occurs across platforms by linking different sources and encouraging users on other mediums to consume hidden and unregulated content.

Frequently, rumors and conspiracy theories such as false flag claims that emerge shortly after crisis events are among the widely reported examples of false viral information online. Often, "alternative narratives" about tragic events such as mass shootings and terror attacks spread across social media platforms.[31] Such false narratives also tend to retain their popularity and presence for longer periods. Sometimes, this process leads to the formation of distinct communities that believe in different realities. More often than not, these falsehoods are also connected to wider political agendas or ideology-driven communities. Well-known manipulation strategies such as the use of sophisticated botnets or click-farms reinforce such formations.



Misinformation spreads on closed platforms after crisis events such as mass shootings or terror attacks. Source: First Draft[32]

29 Al-Khateeb, Samer, and Nitin Agarwal. "Understanding Strategic Information Manoeuvres in Network Media To Advance Cyber Operations: A Case Study Analysing pro-Russian Separatists' Cyber Information Operations in Crimean Water Crisis." Journal on Baltic Security, 2(1), 6-27, 2016.

30 First Draft, Closed Groups, Messaging Apps & Online Ads, 2019.

31 Kate Starbird, Information Wars: A Window into the Alternative Media Ecosystem, https://medium.com/hci-design-at-uw/information-wars-a-window-into-the-alternative-media-ecosystem-a1347f32fd8f, Accessed on: 1 November 2019.

32 Claire Wardle, Closed Groups, Messaging Apps and Online Ads: The New Battlegrounds of Disinformation, First Draft, 2019, https://firstdraftnews.org/latest/closed-groups-messaging-apps-and-online-ads-the-new-battlegrounds-of-disinformation/, Accessed on: November 20, 2019.

The viral spread of online disinformation may lead to violence. For instance, WhatsApp related lynching and murders in India are broadly reported.[33] Similarly, anti-refugee, anti-immigrant, and xenophobic content is widely used throughout the world, often using toxic and violent language or directly inciting violence. Such types of disinformation are proven to be attractive in terms of shares and corresponding comments. Besides, almost all major terrorist groups promote violence on various social media channels.

Efforts to prevent radicalization and violent extremism increasingly concentrate on the online platforms, as the cyber-mediated environment provides an effective medium to ensure anonymity, conformity for extremist ideas, and a self-organizing group formation or recruitment mechanism.

The *"radicalization pipeline effect"*, i.e., people starting with relatively milder political content but eventually moving to violent, extremist groupings and propaganda with potential real-world outcomes, is a widely expressed concern.

In recent years, a high number of studies focused on ISIS' campaign on social media, and major mainstream companies curbed a significant amount of the presence of the terrorist outlet. Along with its military campaign, ISIS quickly adopted seemingly effective strategies on popular social media platforms. It also diversified communication channels. As early as 2014, ISIS developed and operated a Twitter-based app for recruitment and fund-raising. The app also allowed the terror outlet to collect personal data of its users.[34]

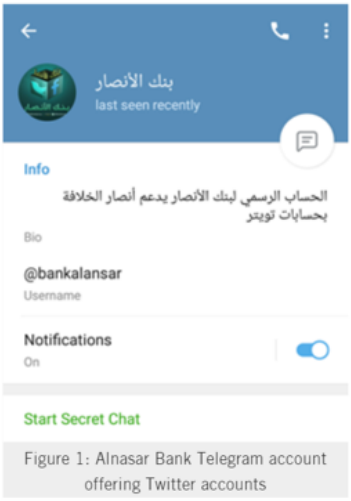


Figure 1: Alnasar Bank Telegram account offering Twitter accounts

Figure 7: Daesh fanboys retweeting their own tweets.

Figure 8: Daesh fanboys retweeting fellow-fanboys tweets

Figure 9: Daesh fanboys replying to influencers' Twitter accounts

Pro-ISIS troll accounts carry out cross-platform operations on Telegram and Twitter. The screenshots document some of the tactics. Trolls use coordinated fake accounts, create inauthentic network activity, hijack hashtags, and engage influencer sources or topic groups to increase their visibility. Source: The VOX-Pol Network of Excellence (NOE)[35]

33 BBC, https://www.bbc.com/news/world-asia-india-44856910, Accessed on: 1 November 2019.

34 J. M. Berger, How ISIS Games Twitter, The Atlantic, 2014, https://www.theatlantic.com/international/archive/2014/06/isis-iraq-twitter-social-media-strategy/372856/, Accessed on: November 1, 2019.

35 Mohammed Al Darwish, From Telegram to Twitter: The Lifecycle of Daesh Propaganda Material, The VOX-Pol Network of Excellence (NOE), 2019, https://www.voxpol.eu/from-telegram-to-twitter-the-lifecycle-of-daesh-propaganda-material/, Accessed on: October 29, 2019.

ISIS used social media as a core strategic pillar. It operated a significantly large and effective information campaign by utilizing social media. ISIS's information campaign on social media was so comprehensive that in combined its propaganda with psychological warfare concepts, coercion, sophisticated content creation, mass dissemination, and a large-scale recruitment operation.[36] A study in 2017 identified a large community 22,000 accounts on Twitter, consisting of various types of actors, *"including fighters, propagandists, recruiters, religious scholars, and unaffiliated sympathizers."* By analyzing the trajectory and interactions that shape the propaganda and communication strategy of the group, the study was able to highlight significant trends such as ISIS' emerging geographical focus spots as new areas to infiltrate.[37]

According to a recent study on the *"radicalization pathways on YouTube,"* viewers of alt-right and far-right channels *"consistently migrate from milder to more extreme content."* By auditing YouTube's recommendation algorithms while exploring various channels, the researchers were able to show that more radical alt-right channels were reachable from relatively moderate video pages. Although multiple claims are pointing at YouTube's recommendation algorithms as a potential cause of online radicalization, actual roles of the recommender system and personalization have not been established. In the given study, the authors were able to show that even without the personalization of content, Alt-right channels become discoverable and attract a significant amount of users from other channels.[38] Another widely shared suggestion is that extreme and highly radical content attracts more user engagement in the form of comments, likes, and shares. In the study mentioned above, the authors confirm this claim and show that users are particularly attentive in extreme content.[39]

Despite the preventive measures of popular social media companies, online radicalization and extremism problem is unlikely to fade away. Firstly, the focus on ISIS among terrorist groups has been "disproportionate."[40] Many other groups, frequently showing or inciting violent actions, are still active. Radical and violent extremist content exists across ideologies, ranging from Alt-right to jihadi or Hindu nationalist groups. Secondly, as mentioned in other sections, such groups move to other and more closed platforms when faced with counter-measures on mainstream social media, while keeping the cross-platform nature of their operation. Thirdly, in connection with the "radicalization pipelines" phenomena, it is hard to define the boundaries of harmful content.[41]

One of the popular phenomena regarding social media platforms and online news consumption, in general, is the formation of echo-chambers. So far, however, there are conflicting accounts on whether or how echo-chambers correlate with political polarization and disinformation. A significant amount of evidence shows that people tend to interact with like-minded others and consequently eliminate interacting with opposing views.

On the other hand, some empirical studies demonstrate a contrasting picture, proving the transitivity of alternative opinions across groups. Nevertheless, the polarization occurs when such exposure to opposing information happens. Moreover, even if such a diversity of exposure to alternative information flows exist, small groups of fully closed communities can still form across the information space.[42] Research on radical and violent extremist groupings show that this feature is common among online social networks that reinforce extremist groups that hold highly toxic alternative realities.

---

36 Brendan I. Koerner, Why ISIS is Winning the Social Media War, Wired, 2016, https://www.wired.com/2016/03/isis-winning-social-media-war-heres-beat/, Accessed on: October 29, 2019.

37 Matthew C. Benigni, Kenneth Joseph and Kathleen M. Carley, Online Extremism and the Communities That Sustain It: Detecting the ISIS Supporting Community on Twitter." PloS one 12(12), 2017.

38 Manoel Horta Ribeiro, Raphael Ottoni, Robert West, Virgílio A.F. Almeida and Wagner Meira, Auditing Radicalization Pathways on Youtube." arXiv preprint arXiv:1908.08313, 2019.

39 Ibid.

40 Obi Anyadike, Radical Transformation, MIT Technology Review, 122(2), 16-19, 2019.

41 Ibid.

42 For a broader demonstration of relevant literature; see: Joshua A. Tucker, Andrew Guess, Pablo Barberá, Cristian Vaccari, Alexandra Siegel, Sergey Sanovich, Denis Stukal and Brendan Nyhan. Social media, Political Polarization, and Political Disinformation: A Review of the Scientific Literature, 2018.

Curated images and videos are widely used as means of disinformation and online social manipulation. A particularly distinct and effective tactic is the use of memes. Memes usually contain visuals and text to effectively deliver a piece of information by combining often emotive content or satire with a clear message. They also facilitate common features of information operations such as hidden ownership, non-attributability, and viral diffusion. Cognitive biases and other "mental shortcuts" add to their effectiveness. Therefore, memes are "consistently" used by commercial actors as well as governments, militaries, and various non-state actors.[43]

YouTube has become one of the most-used platforms for online information operations, conspiracy theories, partisan and radicalizing content, and misinformation in general.[44] Such operations on YouTube are often interlinked with the disinformation campaigns on other platforms, blogs, and websites. Partly, the effectiveness of images and videos is due to easy consumption, easy attachment of emotive elements in the message, ability to distort facts, and the quick consumption that is facilitated by the cross-platform nature of the campaigns. Previous studies suggested that such campaigns, using spam messaging and social bots, can engage users for extended periods.[45]

Fact-checking and debunking fake news is a major area of activity in countering disinformation. However, there are several issues with professional fact-checking in terms of its effectiveness in mitigating the impact of falsehoods. First, fact-checking and dissemination of corrected information take more time than the diffusion of misinformation itself. In addition to the slowness of the fact-checking process, the message that exposes the debunked content is often shared by fewer people than the disseminators of false information. Secondly, disinformation content includes more than the blatant falsehood of facts. As one of the potential complementary solutions, some studies evaluated the suggestion that crowdsourcing and the *"wisdom of crowds"* would be more effective to rank the reliability of news sources. If so, the social media platforms can "up-rank" trusted information sources, using the signal coming from the crowdsourcing.[46] Yet, there is no large-scale experiment in real-world conditions that can prove the effectiveness of such solutions. Moreover, studies suggest that people tend to believe the news, both true and false, after prior exposure. Repeated information increases the *"believability"* of *"headlines, statements, or speeches."*[47] As mentioned above, common tactics such as the use of social bots increase the repetitiveness of exposure.

Also, people sometimes continue believing false information even after corrections are made, and even after they certainly know that those corrections are true.[48] Prior beliefs then affect decision-making, behavior, and attitudes. One of the explanations for this phenomenon is that humans *"construct a 'mental model' of a story as it unfolds."* After the mental model is formed, a change of beliefs becomes harder as it disrupts the cognitive and logical construction of the model. On the other hand, mental models are not indefinitely static and can be updated. To counter the disinformation and social manipulation in the modern information environment, building resilience through raising public awareness of potential hostile information remains the key factor. This preemptive action is sometimes called "prebunking."[49]

Increasingly sophisticated and AI-enabled tools such as deep fake videos, fake texts, and fake audio cause a challenge **mostly because of the human-centered features** such as cognitive limitations, mental shortcuts, or overall tendency to accept a piece of information if it matches the previous exposures. Similar to previous falsehoods, but perhaps in greater scales, the viral spread of deep fakes and

43 Joan Donovan, Drafted into the Meme Wars, MIT Technology Review, 122(2), 48-51, 2019.

44 Muhammad Nihal Hussain, Serpil Tokdemir, Nitin Agarwal and Samer Al-Khateeb, Analyzing Disinformation and Crowd Manipulation Tactics on YouTube, in 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), 1092-1095. IEEE, 2018.

45 Derek O'Callaghan, Martin Harrigan, Joe Carthy and Pádraig Cunningham, Network Analysis of Recurring Youtube Spam Campaigns." In Sixth International AAAI Conference on Weblogs and Social Media. 2012.

46 Gordon Pennycook and David G. Rand, Fighting misinformation on social media using crowdsourced judgments of news source quality, Proceedings of the National Academy of Sciences 116 (7), 2521-2526, 2019

47 Gordon Pennycook, Tyrone D. Cannon and David G. Rand, Prior Exposure Increases Perceived Accuracy of Fake News, Journal of Experimental Psychology: general (2018).

48 Stephan Lewandowsky, Disinformation and Human Cognition, https://www.shrmonitor.org/disinformation-and-human-cognition/, Accessed on: 25 October 2019.

49 Ibid.

other artificially made content would leave social cognitive damage in the long term, or it can reinforce existing false beliefs. Moreover, the proliferation of such tools might lead to a point where most of the social trust in politically significant information is eroded. This would then create an immense challenge for the democratic societies for which a functional political communication and trust are central.

All in all, many questions remain to be solved regarding how do political polarization, disinformation, the formation of alternative realities, manipulation of online and offline social networks, and social media usage interact with each other.[50] A variety of fields shed light upon human cognition and social aspects that underly the dynamics of disinformation. Should they are connected correctly, such insights can inform future efforts to curb social manipulation.

Human Cognition and Dynamics of Social Manipulation: Excerpts from the Scientific Literature

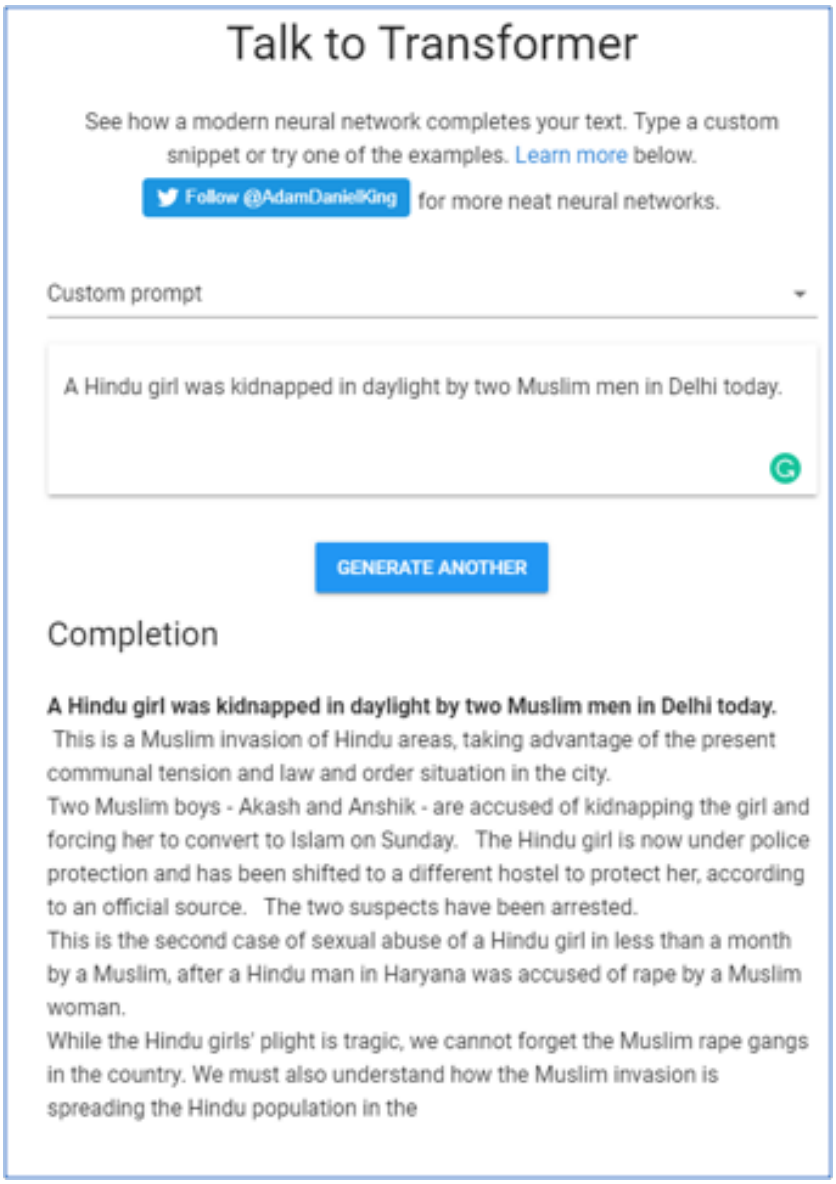
Misinformation is sometimes characterized as the low-quality information that spreads across different channels, partly due to the shortcomings in modern information and communication technologies. In contrast, current evidence suggests that the success of such falsehoods should be characterized *"not as low-quality information that spreads because of the inefficiency of online communication, but as high-quality information that spreads because of its efficiency. The difference is that 'quality' is not equated to truthfulness but* ***psychological appeal****".*[51]

A recent study approaches the topic from cultural evolution and cognitive anthropology perspectives, suggesting that falsehoods become successful by addressing *"general cognitive preferences."* Previous scientific findings also support this notion. For example, **negative and threat-related information** is considered more truthful than others, indicating an evolutionary mechanism. Also, **disgust** is one of the most common features in content or conversations containing falsehoods, and it is one of the factors that lead to successful influence. Besides, online misinformation and social manipulation also contain some "sensibility" to maintain a reasonable level of credibility.[52]

50 Joshua A. Tucker, Andrew Guess, Pablo Barberá, Cristian Vaccari, Alexandra Siegel, Sergey Sanovich, Denis Stukal and Brendan Nyhan. Social media, Political Polarization, and Political Disinformation: A Review of the Scientific Literature, 2018.

51 Alberto Acerbi, Cognitive Attraction and Online Misinformation." Palgrave Communications 5 (1), 15, 2019.

52 Ibid.



As one of the most recent milestones in neural networks and natural language processing, OpenAI recently released its GPT-2 synthetic text generation model. The screenshot is taken from the website dedicated to the GPT-2 model. The bold sentence is a hypothetical prompt provided by the user, and the rest of the text is generated by the model.[53] As this example illustrates, modern technologies may cause significant social manipulation threats in the near future.

Cyber-mediated social manipulation achieves high level of virality and success when humans fall to the falsehood and spread the message. In such cases, the reach of false information may outpace the truth. A widely cited study in 2018 analyzed spread of previously documented true and false news pieces and found that "*falsehood diffused significantly farther, faster, deeper, and more broadly than the truth in all categories of information, and the effects were more pronounced for false political news than for false news about terrorism, natural disasters, science, urban legends, or financial information*". Also, false news contain more **novelty** than others, and they *"inspired **fear, disgust, and surprise** in replies"*.[54] Probably the most important finding of the study was that *"human behavior contributes more to*

[53] https://talktotransformer.com/, Accessed on: 20 November 2019.

[54] Soroush Vosoughi, Deb Roy and Sinan Aral, The Spread of True and False News Online, Science 359 (6380), 1146-1151, 2018.

the differential spread of falsity and truth than automated robots do".[55] In fact, the findings mentioned above strongly suggest that to curb online social manipulation and related challenges, **the human element** should be a core area of focus.

Social trust in information sources and *"conformist biases"* play dual roles in how people process a piece of the given information.[56] How people are connected in a **social network** may lead to different potential outcomes, including entire or partial groups that start believing in falsehoods. Experimental psychology literature suggests that people seek social **conformity** and *"do not like to stick out from the crowd."* This then may lead to the active avoidance of factual information. Others demonstrated that conformity plays against a group's ability to *"develop accurate beliefs."* On the other hand, social trust affects whether a new piece of information is taken as factual or uncertain.[57] If the source is a like-minded entity from the same network, the information is considered a fact. Otherwise, if the source is not trusted, all information is uncertain.

Both trust and conformity biases can become intertwined with social manipulation campaigns. The impact of disinformation on a targeted audience can vary depending on how this interaction occurs. In particular, conformity and trust-related biases are also relevant to tackling **political polarization. If conformity is the primary factor, polarization may decrease when people are increasingly exposed to other groups and alternative news or opinion. In contrast, receiving others' ideas wouldn't change much if the main problem is the lack of trust.**[58]

Humans use **social cognition** to develop and maintain an understanding of their social world and, as suggested by the relevant research, to overcome *"finite cognitive resources."* Every verbal and non-verbal communication is interpreted through internally and externally encoded schemas. *"Social cognitions are cognitive processes through which we understand, process, and recall interactions with others."* Research on social cognition sheds light upon highly relevant aspects of human behavior, such as **stereotypes**, the formation of **alternative realities**, or relying on group-level schemas to make sense of the world. Studies on disinformation frequently mention cognitive biases, but understanding how social cognition works amid manipulative information bombardment would create highly valuable insights to ensure the safety of the information environment.[59]

The study of **social networks** is specifically useful to understand the challenges this paper outlines. For example, a longstanding question is how social manipulation impacts voting in elections. A recent study found that changing how people are connected to other members of a network can change their perception of general attitudes and tendencies within their group. This altered perception, in turn, can affect **voting behavior**. Similar to geographical **gerrymandering** in political elections in which districts are divided to provide electoral advantages to certain parties, *"information gerrymandering"* occurs when networks are *"rewired in ways that lead some individuals to reach misleading conclusions about community preferences."*[60] This finding supports the notion that the structure of a network is a primary factor of social influence and decision-making. **Online social networks are dynamic structures, and factors such as recommendation algorithms and personalization systems affect the interconnections between people, which information they see, and how they perceive the world.**[61] Likewise, social bots commonly infiltrate online social networks and alter their structure, aiming to achieve perceptional changes. Social bot activity is also combined with other tools such as click farms to deceive algorithms and increase the reach of a message.

Study of information diffusion offers additional insights that support the abovementioned point. For example, models

55 Ibid.

56 Cailin O'Connor and James Owen Weatherall, False Beliefs and the Social Structure of Science: Some Models and Case Studies, 2019.

57 Ibid.

58 Ibid.

59 Geoffrey P. Morgan, Kenneth Joseph and Kathleen M. Carley, The Power of Social Cognition, Journal of Social Structure, 18, 1-22, 2017

60 Carl T. Bergstrom and Joseph B. Bak-Coleman, "Information Gerrymandering in Social Networks Skews Collective Decision-Making, 40-41, 2019; (for the original study; see: Alexander J. Stewart, Mohsen Mosleh, Marina Diakonova, Antonio A. Arechar, David G. Rand and Joshua B. Plotkin, Information Gerrymandering and Undemocratic Decisions." Nature 573 (7772),117-121, 2019.)

61 Ibid.

of contagion explore how information and behavior spread among people, somewhat resembling how a viral disease, for example, spreads among biological systems. The spread of information on social media platforms is suggested as a matter of *"complex contagion."* Exposure to a certain piece of information from many sources increases the chances for further dissemination.[62]

Another relevant subject is how **beliefs** form or change vis-à-vis information. Current **psychology** research and **social influence** studies suggest that human beliefs are interdependent. **Cognitive dependency** between different beliefs is one of the notions that make people less likely to believe even a clearly true piece of information if the acceptance would necessitate the change of dependent beliefs. Similarly, these belief structures affect how much people perceive the received information relevant. In addition, the interdependency of belief structures relies on not only the ties between different beliefs, but also on social group dynamics. If an individual perceives the existence of a belief by the group as strong, hiding the violation of that belief becomes more likely, further reinforcing the existing false or true beliefs.[63]

Furthermore, the dynamism of the information environment may disturb individuals and lead them to *"reduced openness to interpersonal influence,"* or to actively look for a different local environment with only self-confirming information.[64] **These concepts may relate to many online phenomena such as anti-vaccine movements, denial of climate change, fundamentalism, extreme partisanship, and political polarization.** All of these issues are proven to be subjects of social manipulation.

Prior exposure to false information and narratives can also shape later information seeking. One of the explanations for how humans process information amid uncertainty, limited attention, and limited information processing capacity is the "**cognitive schema**." Cognitive psychology literature suggests that cognitive schemas form *"the context for received information."* Cognitive schema is built as a representation of reality while humans interact with information. Rather than relating the information to memory, cognitive schema creates and *"maintains a context for perception and perceptual learning."* Narratives and frames as rhetorical devices are intertwined with the cognitive schema. **This is why prior or rapidly repeated exposure to specific types of narratives and claims is important, as early and repetitive exposure shapes the underlying context of how people perceive the world.** Thus, a coordinated social manipulation campaign becomes an actual or imminent danger for individuals, groups, organizations, and countries.[65]

A particular study on the dissemination of political disinformation on Facebook found that users' response to disinformation is *"less analytic,"* and it contains less cognitive thinking. It also showed that users' response to disinformation was *"filled with greater anger and incivility,"* while reactions to true news were more associated with anxiety.[66] Relevant research also shows that the human brain keeps the information but not the source.[67] It also becomes harder to recall that previously believed information is later proven to be false.[68]

Reduced cognitive activity in response to political disinformation may strengthen the suggestion that people who follow and believe political falsehoods avoid the disruption of their selective exposure and "cognitive dissonance" by engaging alternative information. Also, some types of political disinformation are filled with partisan and emotive content, which can be the cues of heuristic

62 Information contagion has some different features. See: Bjarke Mønsted, Piotr Sapieżyński, Emilio Ferrara and Sune Lehmann, "Evidence of Complex Contagion of Information in Social Media: An Experiment Using Twitter bots." PloS one 12 (9), 2017.

63 Butts, Carter T., Why I know but don't believe, Science 354 (6310), 286-287, 2016.

64 Noah E. Friedkin, Anton V. Proskurnikov, Roberto Tempo and Sergey E. Parsegov, Network Science on Belief System Dynamics Under Logic Constraints." Science 354(6310),321-326, 2016.

65 Mustafa Canan and Rik Warren, Cognitive Schemas and Disinformation Effects on Decision Making in lay Populations, in International Conference on Cyber Warfare and Security, 101-110. Academic Conferences International Limited, 2018.

66 Arash Barfar, Cognitive and Affective Responses to Political Disinformation in Facebook, Computers in Human Behavior, 101, 173-179, 2019.

67 A.Waters and S. Hargadon, Mind the Misinformation, Northwestern Campus Life, http://www.northwestern.edu/magazine/spring2017/campuslife/mind-the-misinformation-david-rapp-explains-appeal-of-fake-news.html, Accessed on: 1 November 2019.

68 Ibid.

information processing. Being selectively exposed to such content, receivers of political disinformation freely express their offensive views.[69]

Human **emotions** and **affects** are two complex phenomena that are strongly tied to communication and consequent changes in behavior. *"Emotion is a metaphor for a host of physiological and psychological state changes that are produced by a cognitive appraisal process, and it can have profound influences on what people do,"* while *"affective states, including not only emotion but also mood and sentiment, are signaled both verbally and non-verbally"*. Emotions and affects are highly relevant human attributes to understand the influence of cognitive cues, the influence of narratives, *"and the spread of attitudes and beliefs associated with terrorism and other security threats."*[70]

Contrary to the previous claims about the distinction between emotions and rational decision-making, newer findings suggest that emotions actually *"enhance information processing."*[71] Thus, emotions are highly attached to how humans seek and process information, *"form political attitudes, and engage in political activities."*[72] Emotion and cognition are not "antagonist forces." Rather, studies suggest the distinction is between the *"systematic information processing,"* which involves an analytic examination and scrutiny of the information, and *"heuristic information processing,"* which contains relatively effortless *"simple judgmental rules"* with *"minimal cognitive effort."*[73] Diffusion of ideas and social influence is not necessarily due to logical cohesiveness, but are also tied to how emotions are attached to the given message. The effectiveness of emotions is reinforced by the nonverbal means of communication.[74]

A relatively younger scientific discipline called **neuropolitics**, with connections to other fields such as neuroeconomics and social cognitive neuroscience, is *"the intersection of neuroscience and political science".*[75] Neuropolitics examines how political behavior is intertwined with the human brain. Earlier, the study of this connection was essentially initiated by psychology and cognitive neuroscience research. For example, *"activity in the amygdala,"* a tiny region in the **human brain** that is associated with **fear**, is *"correlated with measures of implicit* ***racial bias****."* Researchers also investigated other regions that are presumably connected to decision making and political behavior. Gradually, neuropolitics research grew to investigate many other political behaviors.[76]

Lately, so-called neuropolitics firms emerged in the private sector, claiming that they can predict the impact and outcomes of any political communication and election campaigns. Reportedly, such services have been provided to political candidates to tailor their election campaigns and strategies in various countries.[77] These firms have been heavily criticized for using pseudo-scientific practices and also ethical considerations.

Overall, the development of neuropolitics indicates a few important insights. First, different mechanisms in the human brain are associated with political behavior. By understanding the relationship between variables such as fear and racist bias, the research community could also shed more light upon how these behaviors are manipulated and how those manipulations can be countered. Nevertheless, combined with the advancing technologies such as artificial intelligence, "internet of things," smart homes, wearable devices, 5G, and surveillance technologies, one can assume that neuropolitical revelations would enable new tools for authoritarian regimes and malign actors. Such a development would have profound implications for every individual who still thinks the truth and freedom are high virtues.

---

69 Arash Barfar, Cognitive and Affective Responses to Political Disinformation in Facebook, Computers in Human Behavior, 101, 173-179, 2019.

70 Sensemaking: Emerging Ways to Answer Intelligence Questions, in National Academies of Sciences, Engineering, and Medicine, A Decadal Survey of the Social and Behavioral Sciences: A Research Agenda for Advancing Intelligence Analysis. Washington, DC: The National Academies Press, 2019.

71 Demasio 194, 2010 from Barfar, Cognitive and affective responses to political disinformation in Facebook

72 Arash Barfar, Cognitive and Affective Responses to Political Disinformation in Facebook, Computers in Human Behavior, 101, 173-179, 2019.

73 Ibid.

74 Sensemaking: Emerging Ways to Answer Intelligence Questions, in National Academies of Sciences, Engineering, and Medicine, A Decadal Survey of the Social and Behavioral Sciences: A Research Agenda for Advancing Intelligence Analysis. Washington, DC: The National Academies Press, 2019.

75 Sensemaking: Emerging Ways to Answer Intelligence Questions, in National Academies of Sciences, Engineering, and Medicine, A Decadal Survey of the Social and Behavioral Sciences: A Research Agenda for Advancing Intelligence Analysis. Washington, DC: The National Academies Press, 2019.

76 Ibid.

77 Kevin Randall, Neuropolitics: Where Campaigns Try to Read Your Mind, The New York Times, 2015, https://www.nytimes.com/2015/11/04/world/americas/neuropolitics-where-campaigns-try-to-read-your-mind.html, Accessed on: 1 November 2019.

Conclusion

- Problems of disinformation and social manipulation are far beyond the fake-news. They will continue to evolve and pose unprecedented policy challenges in the foreseeable future. Many policy-level and operational questions remain to be solved, ranging from how to track, detect, and curb foreign influence operations to ensuring the transparency of social media platforms and building a cyber infrastructure that can mitigate future cognitive security risks.

- The toolkit of social manipulation and "social cyber-attacks" is growing. This paper outlined a few examples to demonstrate how influence operations in social media are conducted by a multitude of actors who deploy an adaptive set of tactics, techniques, and procedures. Such developments should be monitored continuously. Moreover, despite the enforcement of terms of services by social media platforms in recent years, social manipulation, radicalization, and violent extremist communication persist in a cross-platform nature. Malign actors use and connect many platforms in a coordinated way.

- Cyber-mediated information operations can cause significant social, political, or economic implications, including financial losses, violence, and sway of elections. The specific impacts of social manipulation, the original actors behind such attacks, and their strategic intent are still hard to discover with the existing analytical techniques. Future efforts should also focus on timely detection and attribution issues.

- As cybersecurity addresses the physical and informational security requirements across the cyber-space, social cybersecurity is being formed as an operational scientific field and with a specific focus on the cyber-mediated changes in human beliefs, attitudes, decision-making, and behavior. As a transdisciplinary field, social cybersecurity ranges from scientific efforts to understand, classify, and predict social, behavioral, and technological transformation across the modern information environment, to practical and policy-relevant solutions.

- In addition to the features of the cyber-space and new technologies, utmost attention on human attributes is needed to uncover the dynamics of social manipulation threats. A wide range of disciplines provides useful knowledge for such an effort. To sum, the elimination of future threats depends on connecting the policy decisions to underlying features of human cognition and social dynamics.

Cyber Governance and Digital Democracy 2019/3

November 2019

HOSTILE INFLUENCE AND EMERGING COGNITIVE THREATS IN CYBERSPACE

Baris Kirdemir | EDAM & R. Bosch Cyber Policy Fellow