

Pre-Print – final version available at:
https://doi.org/10.1007/978-94-024-1202-4_33-1

Disguised propaganda from digital to social media

Johan Farkas

Johan.farkas@mau.se

Malmö University

School of Arts and Communication

Christina Neumayer

chne@itu.dk

IT University of Copenhagen

Digital Design Department

This is a pre-print of a chapter published in the *Second International Handbook of Internet Research* (eds. Jeremy Hunsinger, Lisbeth Klastrup & Matthew M. Allen). The final version is available online at https://doi.org/10.1007/978-94-024-1202-4_33-1

Abstract

Disguised propaganda and political deception in digital media have been studied since the early days of the World Wide Web. At the intersection of internet research and propaganda studies, this chapter explores disguised propaganda on websites and social media platforms. Based on a discussion of key concepts and terminology, the chapter outlines how new modes of deception and source obfuscation emerge in digital and social media environments, and how this development complicates existing conceptual and epistemological frameworks in propaganda studies. The chapter concludes by arguing that contemporary challenges of detecting and countering disguised propaganda can only be resolved, if social media companies are held accountable and provide the necessary support for user contestation.

Keywords: disguised propaganda, manipulation, disinformation, fake news, deception, social media.

Introduction

@cj_panirman: RT @realDonaldTrump: Time to #DrainTheSwamp in Washington, D.C. and VOTE #TrumpPence16 on 11/8/2016. Together, we will
MAKE AMERICA SAFE ...

@natespuewell: #NeverTrump Those fake, nonsense polls are actually real, good polls, Trump's spokesman insists — Campaign of lies

<https://t.co/Mvja0PPeaH> (Bessi and Ferrara 2016)

At first glance, the above quotes from Twitter during the 2016 U.S. elections come off as deriving from citizens supporting the Republican candidate, Donald Trump and the Democratic candidate, Hillary Clinton respectively. In fact, they were both produced by *social bots*, software-driven social media profiles created to give a deceptive impression of public support (Bessi and Ferrara 2016). Since the dawn of the World Wide Web, political groups and activists have appropriated digital platforms to further partisan agendas based on a range of dissemination tactics (Milan 2013). With the evolution of digital technologies and platforms, communication strategies have developed concurrently, as political actors have iteratively sought to increase their influence. This chapter examines the evolution of disguised propaganda in digital media, relying on identity deception to manipulate online users.

With the development of digital platforms from websites to social media environments, new modes of propaganda have emerged. Situated at the intersection of propaganda studies and internet research, this chapter provides an overview of the development of disguised propaganda from digital to social media. First, the chapter outlines the historical relationship between propaganda and mass media technologies, as conceptualized within the field of propaganda studies. Second, the chapter presents a working definition of disguised propaganda, including the sub-categories of *obfuscated* and *impersonated* propaganda (also defined as *grey* and *black propaganda*). Third, the chapter examines the use of disguised propaganda on relatively static websites, followed by an inquiry into disguised propaganda on social media platforms. Finally, the chapter discusses contemporary and future problems of resistance against disguised propaganda.

Propaganda and media

Although the roots and etymology of propaganda predates electronic communication by several centuries (Auerbach and Castronovo 2013), scholarly engagement with the topic has historically been inseparable from the rise of mass media technologies in the 20th century. Accordingly, the field of propaganda studies has traditionally defined propaganda as intimately connected to media channels such as radio, television, film, and newspapers. In his book *Public Opinion*, Walter Lippmann (1946) referred to “the

manufacture of consent” as “capable of great refinements” (98), a process of manipulation open to anyone who understands and can control it. Herman and Chomsky (1988) in their influential work *Manufacturing Consent* define propaganda as phenomena that “require the collaboration of the mass media” (Herman and Chomsky 1988, 33). Similarly, Ellul argued that propaganda “cannot exist without using these mass media” (Ellul 1965, 9). Later, Cunningham (2002) wrote that it would be “problematic for us to read anything like modern and contemporary propaganda back into periods before the emergence of mass media and mass communication” (17-18). Propaganda, as manifested in the 20th century, is perceived as a distinctly modern phenomenon interwoven with channels of mass communication.

This points to a second characteristic of propaganda, as defined within the field of propaganda studies, namely the *source* of propaganda. Many scholars define propaganda as *de facto* propagated by large organizations such as political and military institutions or commercial corporations (see Ellul 1965, Herman and Chomsky 1988, Sproule 1994). As Sproule (1994) argues:

Propaganda represents the work of *large organizations or groups* to win over the public for special interests through a *massive orchestration* of attractive conclusions packaged to conceal both their persuasive purpose and lack of sound supporting reasons.

(Sproule 1994, 8, added emphasis).

From this perspective, propaganda encompasses mass-mediated manipulation organized on a grand scale to persuade a public. The goal of such persuasion is typically “closely attuned to elite interests” (Herman and Chomsky 1988, 32). Not all scholars, however, exclusively attribute propaganda to elite groups. This is central to this chapter, as digital media’s decentralized mode of content production challenges the perception that propaganda only derives from centralized sources. As a conceptual vocabulary to distinguish between hierarchical and non-hierarchical propaganda, Ellul (1965) proposes the concepts of *vertical* and *horizontal* propaganda.

As part of a rigorous typology, Ellul (1965) introduces the concepts of vertical and horizontal propaganda to distinguish between propaganda from societal elites and from

small citizen groups. Accordingly, Ellul (1965) argues that not all propaganda is connected to political, military, or commercial organizations, although it is “by far the most widespread” (80). To Ellul (1965), vertical propaganda is characterized by originating from elites who rely on mass media to persuade an audience into submission and action. One-to-many communication channels are vital in this regard, as they are means of mass mobilization of crowds to do the bidding of the source (e.g. the government, party leader, general, or company).

Vertical propaganda is particularly effective in *propaganda of agitation*, which is created to mobilize crowds against a portrayed enemy, a “source of all misery” (Ellul 1965, 75). Hitler’s campaigns against the Jews and Lenin’s campaigns against the Kulaks are both examples of such propaganda. In practice, Ellul (1965) argues, agitational campaigns always derive from societal elites. Yet, propaganda of agitation can be effective in getting the audience to take ownership of the constructed narrative, amplifying and extending it. If successful, agitational propaganda does therefore not necessarily rely on a continuous orchestration of mass media, as “each person seized by it” can in turn become their own “propagandist” (Ellul 1965, 74).

Conversely to *vertical* and *agitational* propaganda, *horizontal* and *integrational* propaganda are aimed at “stabilizing the social body, at unifying and reinforcing it” (Ellul 1965, 75). These forms of propaganda can originate from both societal elites, such as governments seeking to stabilize societies during political crises, and from citizens. *Horizontal propaganda* relies on small, autonomous groups cooperating based on a common ideology. It is distinct by deriving from inside the population and “not from the top” (Ellul 1965, 81). According to Ellul (1965), this form of propaganda is a rare phenomenon, non-existent before the 20th century. The decentralized propaganda of Mao’s China is highlighted as an example. Unlike vertical propaganda, which “needs the huge apparatus of the mass media of communication” (Ellul 1965, 82), horizontal propaganda only relies on a “huge organization of people” (82). Media control, in other words, is inseparable from vertical propaganda, but is not similarly fundamental for horizontal propaganda. Table 1 presents an overview of the conceptual differences between the two.

<Insert **Table 1** - Characteristics of vertical and horizontal propaganda >

Propaganda and digital media

Within the last decade, scholars in propaganda studies have argued that the rise of digital media complicates existing theoretical understandings of propaganda, specifically the relationship between propaganda, large-scale organizations, and mass media (Jowett and O'Donnell 2012; Auerbach and Castronovo 2013). Models of propaganda that illustrate how media conglomerates and governments together manufacture systemic biases have somewhat lost their pertinence (Auerbach and Castronovo 2013). The fundamental notion that propaganda relies on centralized control of mass media by large-scale organizations does not seem to encapsulate digital spaces in which propaganda can potentially derive from a multitude of sources. This development signifies a transformation of how scholars should analyze and approach propaganda:

The Internet is now becoming an increasingly important source of information in our society... The potential for propaganda in such a climate is *infinite*. *Anyone* can spread a message, true or false, or manipulate information or even alter a picture to suit his or her own ends.

(Jowett & O'Donnell 2012, 160, added emphasis)

In online environments millions of users operate with crosscutting motives and goals, which challenges and alters the analysts' task of identifying coordinated propaganda campaigns and their underlying political agendas. Clear-cut conceptual boundaries, such as that between vertical and horizontal propaganda, seem to be reshaped into complex continuums going from propaganda at the individual level to that of nation states and multinational conglomerates. Digital media platforms introduce new modalities of propaganda, such as the use of social bots (Shao et al. 2017) and state-organized 'troll armies' (Aro 2016). Taken together, these changes could seem to represent the end, or at least a new beginning, for the field of propaganda studies, but this would be an oversimplified conclusion. Although the internet challenges existing conceptualizations of propaganda, scholars should not be "lulled into thinking that information is now open and free for all" (Auerbach and Castronovo 2013, 12). Digital media's potential for decentralized communication, in other words, should not be equated to a fundamental democratization of information, transcending existing power relations and control.

Herman and Chomsky (2008) argue that the rise of digital media represents a vital new means of communication for political movements across the globe. Yet, the internet should not be seen as a fundamentally democratizing force, destabilizing societal elites and their ability to exercise control through mass orchestrated propaganda. As with all new communication technologies, Herman and Chomsky argue, the internet will first and foremost serve elite and corporate interests. Consequently, the internet functions as means of control to those already in power more than it represents “an instrument of mass communication for those lacking brand names, an already existing audience, and/or large resources” (Herman and Chomsky 2008). Rather than re-inventing the wheel, we need to build upon conceptualizations and terminology of propaganda studies to understand the consequences of disguised propaganda in digital and social media.

Disguised propaganda: Obfuscated and impersonated forms

Propaganda has been defined as the “deliberate, systematic attempt to shape perceptions, manipulate cognitions, and direct behavior to achieve a response that furthers the desired intent of the propagandist” (Jowett and O’Donnell 2012, 7). Drawing on this definition, *disguised propaganda* can be defined as the *deliberate use of disguised sources to manipulate and shape perceptions to achieve a desired outcome*. Within the field of propaganda studies, this specific type of manipulation has also been labeled covert (Ellul 1965; Sproule 1994; Linebarger 2010), clandestine (Soley and Nichols 1987), or concealed propaganda (Jowett and O’Donnell 2012). These terms all point to a specific form of deceptive manipulation, relying on a tactical blurring and misattribution of sources. Drawing on Hancock (2012), disguised propaganda can be defined as a form of *identity-based deception*, which stands in contrast to *message-based deception* (in which content is manipulated, rather than its source). As such, disguised propaganda relies on the manipulation of sources, but not necessarily on the falsification of presented messages (although the two are often interconnected).

Overall, disguised propaganda can be divided into two sub-categories: *grey* and *black propaganda* (Jowett and O’Donnell 2012). In grey propaganda, sources are deliberately obfuscated, making it either impossible or difficult to identify the propagandist hiding underneath (Sproule 1994). In black propaganda, disseminated messages are attributed to a false source, which is “presented by the propagandizer as coming from

a source inside the propagandized” (Becker 1949). Black propaganda, in other words, relies on deceiving an audience into believing that a distributed message derives from an opposing source to the actual one (e.g. an ally rather than an enemy). Both grey and black propaganda stand in contrast to *white propaganda*, which encompasses manipulation in which the actual source is known and visible. The conceptual vocabulary of white, grey, and black propaganda has been used extensively within the field of propaganda studies (Becker 1949; Ellul 1965; Daniels 2009a; Soley and Nichols 1987; Jowett and O’Donnell 2012). However, as Daniels (2009a) argues, these concepts have a substantial downside due to their unfortunate racial connotations. Building on this critique, this chapter proposes the terms *identifiable* (white), *obfuscated* (grey), and *impersonated* (black) propaganda as a working alternative to the long-standing, yet problematic terms.

Forms of disguised propaganda represent an innately challenging object of investigation due to the difficulty of untangling concealed sources and intentions (Jowett and O’Donnell 2012). In some cases, manipulated sources can only be studied retrospectively after historical documents surface (Soley and Nichols 1987). Yet, analysts can advantageously deploy a number of investigative strategies to determine the hidden identity of the propagandist. One such strategy relies on analyzing “the apparent ideology, purpose, and context of the propaganda message. The analyst can then ask, who or what has the most to gain from this?” (Jowett and O’Donnell 2012, 293). If analysts can establish the intended outcome of disguised propaganda, this will point to the underlying source. Studying the context and effects of propaganda are key in this regard. To Jowett and O’Donnell (2012), the hidden source will typically be “an institution or organization, with the propagandist as its leader or agent” (293). Using the conceptual vocabulary of Ellul (1965), disguised propaganda is thus first and foremost conceptualized as a form of vertical propaganda.

An early example of impersonated propaganda of agitation in mass media is *The Protocols of the Elders of Zion*, which was written by the Czar Nicholas II’s secret police in 1903 and distributed through Russian newspapers (Jowett and O’Donnell 2012). On the surface, the text reveals a devious Jewish plot for world domination conceived by Jewish representatives at a secret congress. In reality, it is a deliberate fraud to promote anti-Semitism. Nonetheless, the text became influential in European politics,

cited by Hitler in his infamous *Mein Kampf* and used in Nazi propaganda (Jowett and O'Donnell 2012). As such, it accompanied the *psychological warfare* of the Second World War.

Disguised propaganda plays a key role in modern-day psychological warfare, which encompasses “the use of propaganda against an enemy together with other operational measures of a military, economic, or political nature” (Linebarger 2010, 40). During both The Second World War and The Cold War, military units and intelligence agencies in countries such as Germany, The United Kingdom, The U.S., and The Soviet Union all orchestrated large-scale disguised propaganda campaigns against their enemies (Becker 1949; Soley and Nichols 1987). Radio represented a particularly powerful medium in this regard, as it enabled fast and widespread dissemination of subversive content into enemy territories (Soley and Nichols 1987). Despite the effectiveness of clandestine radio, such propaganda posed great challenges to its creators, as it required both an elaborate orchestration of media technologies as well as “operatives thoroughly acquainted with every relevant aspect of the society and culture in question” (Becker 1949, 224).

The development of digital media technologies complicates the fundamental notion that disguised propaganda *de facto* derives from large-scale organizations through one-to-many communication channels. With digital media's decentralized mode of content production and propagation, the number of potential sources has risen dramatically. This complicates existing analytical frameworks for identifying and analyzing sources of disguised propaganda. Nonetheless, the prominence of digital media should not be seen as the end of large-scale propaganda orchestration (see Auerbach and Castronovo 2013; Herman and Chomsky 2008).

Disguised propaganda in digital media

Identity deception and disinformation have been studied since the early days of the World Wide Web. Some of the first to discuss the risks of online manipulation were scholars from information science, studying the internet's role as an educational tool (see SantaVicca 1994; Tate and Alexander 1996). In other disciplines, scholars studied identity deception in Usenet groups (see Donath 1998; Dahlberg 2001) and the potential risk of the internet becoming a “disinformation superhighway” (Floridi 1996, 509).

Steering away from a techno-determinism, Floridi (1996) argued that, although “[t]echnology sharpens the problems [of disinformation]...the fundamental questions remain human and social” (513). Deception and manipulation, in other words, might take new forms online, yet the underlying roots and causes would remain the same.

Dahlberg (2001) was one of the first to argue that deception represented a potential hindrance for the internet to ever become a space of democratic deliberation:

Many discussion groups, including those dedicated to 'serious' political issues, face the problem postings aimed to misinform, embarrass, self-promote, provoke, gossip, trivialize, and so on... Verifiable online evidence is often hardest to come by in cases where support for claims is most crucial... These verification problems can inhibit online interactions from realizing the deliberative conception where only 'the force of better argument' decides outcomes.

(Dahlberg 2001, 19-20)

Donath (1998) presents a similar argument, stating that deception was both common and harmful in digital media environments. To Donath (1998), limited identity cues online made deception particularly treacherous, as information “is more likely to be believed when offered by one who is perceived to be an expert” (Donath 1998, 31). Any user could deceive others by disguising their identity behind profiles or websites claiming to be trustworthy and authoritative. Accessibility and affordability of content production made deception much easier, as “documents, photographic evidence, and whole organizations can be readily fabricated” (Dahlberg 2001, 19). An early example of such deception strategies can be found on martinlutherking.org.

Martinlutherking.org, which was launched in 1999 (Thomson 2011), is a website deliberately constructed to promote white supremacy based on a difficult-to-identify source. At first glance, the website appears to be educational and scientific by claiming to provide ‘A True Historical Examination’ of Martin Luther King. Nonetheless, the site (which is still active at the time of writing this chapter) one-sidedly portrays Dr. King as a rapist, communist, women-beater, and sexual deviant (Daniels 2009a, 2009b). Drawing on the conceptual works from propaganda studies (as discussed in

the previous sections), martinlutherking.org is a case of obfuscated propaganda, as the site does not clearly disclose its authorship. In 2008, Daniels conducted a pilot study in which she asked adolescent internet users to search for Martin Luther King on Google and find a suitable website for a school paper (Daniels 2008). The result of the study showed that numerous adolescents – including experienced internet users – found martinlutherking.org and were unable to identify the disguised authorship. A majority of participants concluded that the website would be a suitable source for a school paper.

Apart from being a case of obfuscated propaganda, martinlutherking.org is also horizontal propaganda, as the website is not created by a large-scale institution, but by an individual, Don Black – a former grand wizard of the Ku Klux Klan who funded the website through donations (Daniels 2009b). As noted, disguised propaganda in radio, newspapers, film, and television has historically been closely connected to large-scale organizations, as these channels required substantial resources. In digital media, this is different, as expenses for buying and maintaining a web domain is minimal in comparison to mass mediated campaigns. This enables small groups or individuals to orchestrate campaigns, including propaganda of agitation. In the typology of Ellul (1965), horizontal propaganda of agitation is non-existent, as individuals cannot orchestrate the necessary media technologies. Following Daniels' (2014), however, digital media enable such propaganda due to low barriers of content production and proliferation. This makes the internet potentially powerful for individuals or groups seeking to further political agendas through manipulative means:

One of the many promises of digital media is that it opens up the possibility for multiple perspectives... If the wonder of the open Internet is that anyone can create and publish content online, it is also simultaneously the distress, as those who intend to deceive create and publish cloaked websites. (Daniels 2014, 151)

Daniels (2009a, 2014) uses the term *cloaked websites* to designate disguised propaganda on websites. Other notable examples of such propaganda promoting white supremacy are *The Institute of Historical Review* (Daniels 2009b; Foxman and Wolf 2013), *American Civil Rights Review* (Daniels 2009b) and *The Occidental Quarterly* (Mihailovic 2015). All these sites disguise their underlying political goals and ideologies. Racism, however, is not the only disguised agenda online. Various websites

have used similar tactics. Teen Breaks, is an example of a disguised, pseudo-scientific website aimed at convincing young pregnant women to renounce abortion (Daniels 2014). Makah.org was an impersonated (or black) propaganda website created by animal rights activists to discredit the Makah Indian Tribe for harvesting whales (Piper 2001). Gwbus.com was a parody website created by left-wing activists during the George W. Bush election campaign to deceive people into thinking it was an official website and to mock Bush (Foot and Schneider 2002). A majority of these websites are *horizontal propaganda*, as they do not derive from societal elites or powerful organizations. Yet, large-scale corporations have also orchestrated disguised campaigns through websites.

Astroturfing refers to persuasion campaigns orchestrated by organizations to give a false impression of public support for or against a specific topic, which serves their agenda (Leiser 2016). The term derives from *AstroTurf*, which is a brand of synthetic turf playing grass, thus highlighting the contrast between orchestrated support and actual grass roots movements (Zhang et al. 2013). Astroturfing is a form of *disguised, vertical propaganda* aimed at creating the impression of horizontal, political support. Notable early examples of astroturfing websites are *Working Families for Wal-Mart*, *The Center for Food and Agricultural Research*, and *Americans for Technology Leadership* (Leiser 2016). These organizations all claimed to represent independent advocate groups, yet were in fact funded and orchestrated by Wal-Mart, Monsanto, and Microsoft respectively. These corporations tactically used the organizations to counter negative attention towards their brands and attack commercial and political opponents. Governments in countries such as China and Russia have in recent years used similar tactics on a much greater scale, relying on paid users to influence political discourse on social media (Tong and Lei 2013; Aro 2016; King et al. 2017).

All in all, digital media have complicated existing conceptualizations of disguised propaganda as de facto deriving from large-scale organizations. Yet, as the above examples highlight, it would be problematic to assume that digital media have erased vertical propaganda. Lines between vertical and horizontal propaganda seem increasingly blurred, as individuals, groups, and powerful organization can all potentially create and orchestrate disguised campaigns within the same online environments. Political and military organizations have systematically sought to take advantage of this

situation, ushering in a new era of propaganda, surveillance, and censorship:

Civilian communication networks, including the Internet, are now fully intertwined with military communications, a situation that has led to networks being retooled for surveillance, control and information warfare. These pressures are also eroding formerly distinct elements of media–public diplomacy–military relations...

(Winseck 2008, 420)

The boundaries between national and global media systems, media producers and consumers, military operations and politics become increasingly fluid online. The following section seeks to examine this development, focusing on the global rise of social media – e.g. Facebook, Twitter, Instagram, Snapchat, and WeChat – and the continued evolution of vertical and horizontal disguised propaganda on these platforms.

Disguised propaganda on social media

Social media are online platforms that enable large-scale proliferation of user-generated content based on many-to-many communication (Castells 2013). However, since all media technologies that support human communication can essentially be considered social, the term social media “obscures the unpleasant truth that ‘social media’ is the takeover of the social by the corporate” (Baym 2015, 1). During the last decade, the number of social media users have grown exponentially, with Facebook reaching two billion users in 2017 (Chaykowski 2017). As such, human interaction – whether in relation to political, cultural, or everyday life – increasingly takes place in social media environments. This development gives rise to new modalities of both vertical and horizontal propaganda, produced by propagandists and validated and amplified by users through comments, likes, re-tweets, and shares.

Corporate social media platforms have a profound influence on social relations, as they not only facilitate interactions, but also actively shape them:

Sociality is not simply "rendered technological" by moving to an online space; rather, coded structures are profoundly altering the nature of our connections, creations, and interactions. Buttons that impose "sharing" and "following" as social values have effects in cultural practices and legal disputes, far beyond platforms

proper.

(van Dijck 2013, 20)

Researchers should consequently approach social media with attentiveness towards the interrelation between technological and social processes. This, however, is a difficult task, as the visibility of *how* social media influence social relations through algorithms and interfaces (and *vice versa*) has not become greater alongside their ubiquity (van Dijck 2013). Following this argument, disguised propaganda on social media should not simply be seen as old wine in new bottles, but as new, socio-technical phenomena arising at the intersection of social relations and digital architectures. In this context, user engagement is central, as social media content spreads through user networks. Studying disguised propaganda on social media thus requires researchers to closely examine the relationship between producers, audiences/distributers, and platform architectures.

In the context of horizontal propaganda, social media have lowered the cost of digital content production. Whereas websites (only) required the purchase and maintenance of a web domain, social network sites (SNSs) are available to anyone with a working computer or smartphone with Wi-Fi access. This has opened up a new venue for individuals and small groups seeking to further agendas through manipulation. Cloaked websites, as a form of *obfuscated propaganda*, typically present content as serious and trustworthy, while concealing the website's authorship (Daniels 2009a). SNSs such as Facebook, which are based on personal profiles and the display of personal networks, are particularly well suited for *impersonated* forms of propaganda. Reliability and trust of a disguised social media profile is created by carefully constructing a false identity and maintaining it through posts that are validated through user comments, likes and shares. A user who 'likes' or 'befriends' a disguised profile or page can thus potentially contribute to both its distribution and validation.

A popular definition of SNSs argues that their key characteristics are the ability to create a public or semi-public profile, make connections known, and view and navigate these connections (boyd and Ellison 2007) These authors later revised the connectivity function by including production of and interaction with streams of user-generated con-

tent (Ellison and boyd 2013). The profile itself can be defined as “a portrait of an individual as an expression of action, a node in a series of groups, and a repository of self- and other-provided data” (Ellison and Boyd 2013, 154). Similarly, conceptualizing social media more generally as “personal media assemblage and archives” (Good 2013, 560) allows us to consider the identity created through SNSs as central for analyzing disguised sources. Combining these characteristics, disguised propaganda on social media is: based on a (cloaked) identity created through a profile or page; a stream of user-generated content; and an identity that is continuously reproduced and negotiated in interactions between posts and comments. Moreover, disguised propaganda on SNSs is rarely permanent, but exists in an interactive process of creation, deletion (due to violations of SNS platforms’ terms of use), and recreation (Farkas et al. 2017).

One example of impersonated propaganda on social media is their use to distribute and amplify racist discourses concerning ethno-cultural minorities. In 2015, anonymous propagandists in Denmark successfully provoked users by constructing fake Muslim identities on Facebook, claiming that Muslims were plotting to kill and rape (non-Muslim) Danes (Farkas et al. 2017). Through 11 Facebook pages, propagandists attracted more than 20.000 comments from Danish Facebook users. The most commented page, which existed for less than four days before Facebook deleted it, attracted more than 10.000 comments. A majority of users who reacted to these cases of impersonated propaganda of agitation expressed aggression towards the pages as well as Muslims in general. Through hateful comments, the pages turned into sites of overt hatred and racism. Due to Facebook’s design (which provides almost unlimited anonymity and security to page owners), the propagandists behind the fake identities were able to remain completely anonymous. This differs from cloaked websites where sources are obfuscated, yet often identifiable at closer inspection (Daniels, 2009a). On social media, it can be impossible to establish with certainty, whether individuals or an organized group created a page and posts, i.e. whether it is vertical or horizontal propaganda.

As stated, propaganda analysts can examine disguised sources and intentions by asking the basic question of “who or what has the most to gain from this?” (Jowett and O’Donnell 2012, 293). This analytical strategy is useful in the context traditional mass media, where disguised propaganda is often orchestrated by a limited number of large-scale

organizations (Ellul 1965; Jowett and O’Donnell 2012). On social media, however, this investigation strategy is challenged. Determining ‘who gains the most’ is difficult, when propaganda can potentially derive from a small partisan group, a large-scale organization, or even a single individual seeking to further an agenda or simply provoking others by *trolling* (Phillips 2012). This raises new epistemological challenges (Schou and Farkas 2016): How can we investigate disguised sources and intentions on social media? How does the credibility of a social media profile or page increase through its likes, shares, and comments? How can users become more critical of information streams produced by aggressive and hateful posts as well as antagonistic reactions? And how can we assess the magnitude and significance of disguised propaganda, when fake profiles can reach thousands of users within days, before social media companies delete them? These questions require urgent scholarly attention.

In the context of vertical propaganda, large-scale organizations take advantage of the decentralized structure of social media by orchestrating far-reaching campaigns that are nonetheless difficult to identify as such. Two vital components in this regard are the use of so-called *troll armies* and *social bots* for social media astroturfing (Benedictus 2016). On social media, astroturfing encompasses the orchestration of user profiles by an organization, such as a government agency or private corporation, to simulate public support or opposition towards a particular topic. This form of disguised manipulation can serve as both propaganda of agitation and integration, as organizations seek to consolidate power through attacks on perceived opponents as well through the manufacturing of widespread support. Astroturfing can also rely on both obfuscated and impersonated sources, as organizations might pay users to post content from their own social media accounts or through networks of fictitious profiles. In practice, these modalities are often interconnected. In China and Russia, government agencies have orchestrated large-scale troll armies, in which people are paid to promote government agendas through social media profiles (Tong and Lei 2013, Benedictus 2016). In China, this has been coined the ‘50-cents party’, as users were rumored to receive 50 cents for each social media post they create in support of the government (Aro 2016).

Large-scale organizations engage in social media astroturfing for a number of reasons. This form of disguised propaganda can potentially have widespread influence on public

opinion and be an effective tool to silence critics through aggressive campaigns (Aro 2016). Astroturfing might also serve to divert public attention from contemporary crises by flooding social media with non-related content. This method, which relies on a “strategic distraction from collective action, grievances, or general negativity” (King et al. 2017), has been used extensively in China, where the government is estimated to orchestrate 448 million social media posts per year (King et al. 2017). This content is first and foremost produced by human laborers (King et al. 2017), yet astroturfing can also rely on social bots.

Social bots are user profiles controlled by software that algorithmically produce and disseminate content. During the 2016 U.S. presidential elections, a large-scale study estimates that close to 20 percent of all content on Twitter concerning the elections was produced by social bots (Bessi and Ferrara 2016). Bots were found on both sides of the political spectrum, although a majority supported the Republican candidate, Donald Trump (Kollanyi et al. 2016). A key application of bots was to disseminate conspiracy theories and disinformation, popularly referred to as ‘fake news’ (Shao et al. 2017). As of the time of writing, the U.S. Congress is investigating whether some of this activity was orchestrated by Russian agencies (Wakabayashi and Shane 2017), but despite the potential political implications of these activities, it remains difficult to identify the sources behind automated propaganda:

Concluding, it is important to stress that... it is impossible to determine who operates such bots. State- and non-state actors, local and foreign governments, political parties, private organizations, and even single individuals with adequate resources... [could] deploy armies of social bots and affect the directions of online political conversation.

(Bessi and Ferrara 2016)

As with all disguised propaganda on social media, platforms’ decentralized mode of content proliferation and potential anonymity provided for propagandists complicate both epistemological boundaries and empirical investigations of vertical and horizontal propaganda. This raises serious and urgent questions of accountability, transparency,

and contestation of disguised propaganda – for scholars, users, policy makers, law enforcement, and (perhaps most importantly) the corporations owning these platforms. So far, social media companies have largely placed the responsibility for countering these phenomena on users. However, this strategy is not a viable solution to contemporary or future democratic consequences posed by disguised propaganda.

Countering disguised propaganda on social media

The vast popularity of social media platforms makes it difficult for companies, such as Facebook and Twitter, to identify and moderate problematic content. As a solution to this challenge, companies construct their policy enforcement principles around user engagement, often deploying commercial content moderators solely when users flag content for violations of company policies (Roberts 2016). Reimagining and reengineering this division of labor is a difficult endeavor:

The huge numbers of members that popular social media sites boast and the vast volume of content these members post make it impossible for the staff of the host companies to pro-actively monitor and edit the contents. As we've seen, the only way content guidelines - in particular, those related to hate speech - can be applied is through the active engagement of real people... Almost inevitable, this task falls mainly on the users of the social media sites.

(Foxman and Wolf 2013, 106)

Identifying hate speech and disguised propaganda has to rely on human judgment, as algorithms cannot adequately analyze cultural contexts of each post (at least not yet). Due to the ubiquity of social media platforms, such human judgment has to derive from users. Following this argument, an encouraging solution to disguised propaganda could seem to be the formation of citizen groups, actively fighting propaganda by reporting fake pages and profiles to social media companies. Promising as this initiative may be, users can only superficially counteract disguised propaganda under current conditions (Farkas and Neumayer 2017).

There are many challenges involved in building alternative spaces to fight disguised propaganda on social media platforms. New forms of digital editing tools make it increasingly difficult to determine, if pictures and videos are manipulated. The decentralized structure of social media platforms makes it difficult to find and contest propaganda before it potentially reaches a wide audience. The biggest challenge, however, is the way in which social media companies place the responsibility for countering propaganda on their users, yet only provide limited and opaque opportunities for them to act. As a result, tactics to manipulate users become increasingly sophisticated, while collective resistance cannot.

The idea of empowerment of crowds acting and creating together has been present in early discourses about social media. Tim O'Reilly coined the term Web 2.0 with one key component being the “wisdom of the crowds” (O'Reilly 2005). For social media companies, crowdsourcing became an effective marketing discourse, in which they present their platforms as spaces of participation, decentralization, spontaneous interaction, and lack of hierarchy – ideas hijacked from the radical left (Žižek 2009). In the case of fighting disguised propaganda, these ideas about social media shift the responsibility to the users. They, however, have to navigate limitations of architectures and policies provided by social media corporations. Instead of empowering activists, “power has partly shifted to the technological mechanisms and algorithmic selections operated by large social media corporations” (Poell and van Dijck 2015, 534)

On Facebook, users are provided only with a ‘report’ button to notify the company of content violations (Farkas and Neumayer 2017). How Facebook processes these reports remains highly opaque. Consequently, users cannot know how or on what grounds Facebook takes action. Even *if* Facebook deletes a profile or page, the creators can typically remain anonymous and continue their work. This makes it incredibly difficult for users or authorities to hold anyone accountable. These challenges also complicate the work of journalists or researchers trying to study the implications of disguised propaganda, as a page or profile might reach thousands of users within days and then disappear without notice. To limit the potential contemporary and future threat of disguised propaganda, users should be able to identify, mobilize, organize, and collectively resist

manipulation much more effectively. Although anonymity can be beneficial for democratic discussion in many ways, it is problematic for counter-action that creators of disguised propaganda can stay completely anonymous and avoid any consequences. For this to change, social media corporations need to be held accountable for countering propaganda on their platforms. In the current situation, crowdsourced user actions mainly seem to serve as a diversion from corporate responsibility and questions of accountability.

Conclusion

Disguised propaganda has undergone a series of profound changes alongside technological developments throughout the 20th and 21st century: from impersonated propaganda in early 1900s newspapers (e.g. *The Protocols of the Elders of Zion*), to clandestine radio during the Second World War, all the way up until present-day social bots and troll armies on social media. Alongside this significant evolution, scholarly contributions, such as analytical and epistemological frameworks, are continuously challenged. As this chapter has made apparent, digital media platforms complicate the fundamental notion of disguised propaganda *de facto* deriving from large-scale organizations. Additionally, digital media challenge existing conceptual boundaries, such as Ellul's (1965) conceptualization of vertical and horizontal propaganda. These technical developments, however, do by no means render these profound conceptual works redundant. Contrarily, revisiting concepts of propaganda studies (such as Ellul 1965); Sproule 1994; and Herman and Chomsky 1988) enables us to explore how disguised propaganda changes in digital and social media, but also to outline their continuity across different media technologies. More scholarly engagement with disguised propaganda on social media is necessary to further develop concepts at the intersection of internet research and propaganda studies. Research in this field should expand methodological, analytical, conceptual, and epistemological frameworks, but also support resistance against disguised propaganda that produces hatred and racism. Scholars should not only strive to understand the development of propaganda, but also challenge and contest manipulation and deception in contemporary and future online spaces.

References:

- Aro J (2016) The cyberspace war: propaganda and trolling as warfare tools. *Eur View* 15:121–132. doi: 10.1007/s12290-016-0395-5
- Auerbach J, Castronovo R (2013) Introduction: Thirteen Propositions about Propaganda. In: Auerbach J, Castronovo R (eds) *The Oxford Handbook of Propaganda Studies*. Oxford University Press, Oxford, pp 1–16
- Baym NK (2015) Social Media and the Struggle for Society. *Soc Media + Soc* 1:205630511558047. doi: 10.1177/2056305115580477
- Becker H (1949) The Nature and Consequences of Black Propaganda. *Am Sociol Assoc* 14:221–235.
- Benedictus L (2016) Invasion of the troll armies: from Russian Trump supporters to Turkish state stooges. *Guard*.
- Bessi A, Ferrara E (2016) Social bots distort the 2016 us presidential election online discussion.
- boyd danah m., Ellison NB (2007) Social Network Sites: Definition, History, and Scholarship. *J Comput Commun* 13:210–230. doi: 10.1111/j.1083-6101.2007.00393.x
- Castells M (2013) *Communication Power*, Second Edi. Oxford University Press, Oxford
- Chaykowski K (2017) Mark Zuckerberg: 2 Billion Users Means Facebook’s “Responsibility Is Expanding.” *Forbes*
- Daniels J (2009a) Cloaked websites: propaganda, cyber-racism and epistemology in the digital era. *New Media Soc* 11:659–683. doi: 10.1177/1461444809105345
- Daniels J (2009b) *Cyber Racism: White Supremacy Online and the New Attack on Civil Rights*. Rowman & Littlefield Publishers Inc., New York
- Daniels J (2014) From Crisis Pregnancy Centers to Teenbreaks.com: Anti-abortion Activism’s Use of Cloaked Websites. In: *Cyberactivism on the Participatory Web*. pp 140–154

- Donath JS (1998) Identity and deception in the virtual community. In: Smith MA, Kollock P (eds) *Communities in Cyberspace*. Routledge, London, pp 22–58
- Ellison NB, Boyd DM (2013) Sociality Through Social Network Sites. In: Dutton WH (ed) *The Oxford Handbook of Internet Studies*. Oxford University Press, Oxford, pp 151–172
- Ellul J (1965) *Propaganda: The Formation of Men’s Attitudes*. Vintage Books, New York
- Farkas J, Neumayer C (2017) “Stop fake hate profiles on Facebook”: Challenges for crowdsourced activism on social media.
- Farkas J, Schou J, Neumayer C (2017) *Cloaked Facebook Pages: Exploring Fake Islamist Propaganda in Social Media*. *New Media Soc*. doi: <https://doi.org/10.1177/1461444817707759>
- Floridi L (1996) Brave.Net.World: the Internet as a disinformation superhighway? *Electron Libr* 14:509–514.
- Foot KA, Schneider SM (2002) Online Action in Campaign 2000: An Exploratory Analysis of the U.S. Political Web Sphere. *J Broadcast Electron Media* 46:222–244. doi: 10.1207/s15506878jobem4602_4
- Foxman AH, Wolf C (2013) *Viral Hate*. Palgrave Macmillan, New York
- Good KD (2013) From scrapbook to Facebook: A history of personal media assemblage and archives. *New Media Soc* 15:557–573. doi: 10.1177/1461444812458432
- Hancock JT (2012) *Digital deception: Why, when and how people lie online*. Oxford University Press, Oxford
- Herman ES, Chomsky N (1988) *Manufacturing Consent: The Political Economy of the Mass Media*. Pantheon Books, New York
- Herman ES, Chomsky N (2008) *Manufacturing Consent: The Political Economy of the Mass Media*. The Bodley Head, London

- Jowett GS, O'Donnell V (2012) Propaganda and Persuasion. SAGE Publications, Los Angeles
- King G, Pan J, Robert ME (2017) How the Chinese Government Fabricates Social Media Posts for Strategic Distraction, Not Engaged Argument. GkingHarvard.edu. doi: 10.1017/S0003055417000144
- Kollanyi B, Howard PN, Woolley SC (2016) Bots and Automation over Twitter during the U.S. Election.
- Leiser M (2016) AstroTurfing, “CyberTurfing” and other online persuasion campaigns. *Eur J Law Technol* 7:1–27.
- Linebarger PMA (2010) Psychological Warfare. Coachwhip Publications, Darke County, Ohio
- Lippmann W (1946) Public opinion (Vol. 1). Transaction Publishers, New Brunswick and London
- Milan S (2013) Social Movements and Their Technologies: Wiring Social Change. Palgrave Macmillan, New York
- O'Reilly T (2005) What is Web 2.0: Design Patterns and Business Models for the Next Generation of Software. <http://www.oreilly.com/pub/a/web2/archive/what-is-web-20.html>. Accessed 6 Oct 2017
- Phillips W (2012) The House That Fox Built: Anonymous, Spectacle, and Cycles of Amplification. *Telev New Media* 14:494–509. doi: 10.1177/1527476412452799
- Piper P (2001) Better Read That Again: Web Hoaxes and Misinformation.
- Poell T, van Dijck J (2015) Social Media and Activist Communication. In: Atton C (ed) *The Routledge companion to alternative and community media*. Routledge, London, pp 527–537
- Roberts ST (2016) Commercial Content Moderation: Digital Laborers' Dirty Work. In: Noble SU, Tynes B (eds) *The Intersectional Internet: Race, Sex, Class and Culture Online*. Peter Lang, New York, pp 147–160

- Santa Vicca EF (1994) The Internet as reference and research tool: a model for educators. *Ref Libr* 19:225–236.
- Schou J, Farkas J (2016) Algorithms, Interfaces, and the Circulation of Information: Interrogating the Epistemological Challenges of Facebook. *KOME - An Int J Pure Commun Inq* 4:36–49. doi: 10.17646/KOME.2016.13
- Shao C, Ciampaglia GL, Varol O, Jul SI (2017) The spread of fake news by social bots. In: *ArXiv*. <https://arxiv.org/abs/1707.07592v2>. Accessed 5 Oct 2017
- Soley LC, Nichols JS (1987) *Clandestine Radio Broadcasting: A Study of Revolutionary and Counterrevolutionary Electronic Communication*. Praeger, New York
- Sproule MJ (1994) *Channels of Propaganda*. EDINFO Press, Bloomington
- Tate M, Alexander J (1996) Teaching critical evaluation skills for World Wide Web resources. *Comput Libr* 16:49–54.
- Thomson K (2011) White Supremacist Site MartinLutherKing.org Marks 12th Anniversary. *Huffingt. Post*
- Tong Y, Lei S (2013) War of position and microblogging in China. *J Contemp China* 22:292–311. doi: 10.1080/10670564.2012.734084
- van Dijck J (2013) *The Culture of Connectivity: A Critical History of Social Media*. Oxford University Press, Oxford
- Wakabayashi D, Shane S (2017) Twitter, With Accounts Linked to Russia, to Face Congress Over Role in Election. In: *New York Times*. <https://www.nytimes.com/2017/09/27/technology/twitter-russia-election.html>.
- Winseck D (2008) Information Operations ‘Blowback’: Communication, Propaganda and Surveillance in the Global War on Terrorism. *Int Commun Gaz* 70:419–441. doi: 10.1177/1748048508096141
- Zhang J, Carpenter D, Ko M (2013) Online astroturfing: A theoretical perspective. *19th Am Conf Inf Syst AMCIS 2013* 4:2559–2565.

Žižek S (2009) Violence: Six sideways reflections. Profile Books, London

Pre-print