Studies in Communication Sciences 18.1 (2018), pp. 69–85

Digital astroturfing in politics: Definition, typology, and countermeasures

Marko Kovic, Zurich Institute of Public Affairs Research (ZIPAR), Switzerland*
Adrian Rauchfleisch, National Taiwan University and Zurich Institute of Public Affairs Research (ZIPAR), Switzerland
Marc Sele, Zurich Institute of Public Affairs Research (ZIPAR), Switzerland
Christian Caspar, Zurich Institute of Public Affairs Research (ZIPAR), Switzerland
*Corresponding author: markokovic@gmail.com

Abstract

In recent years, several instances of political actors who created fake grassroots activity on the Internet have been uncovered. We propose to call such fake online grassroots activity digital astroturfing, and we define it as a form of manufactured, deceptive and strategic top-down activity on the Internet initiated by political actors that mimics bottom-up activity by autonomous individuals. The goal of this paper is to lay out a conceptual map of the phenomenon of digital astroturfing in politics. To that end, we introduce, first, a typology of digital astroturfing according to three dimensions (target, actor type, goals), and, second, the concept of digital astroturfing repertoires, the possible combinations of tools, venues and actions used in digital astroturfing efforts. Furthermore, we explore possible restrictive and incentivizing countermeasures against digital astroturfing. Finally, we discuss prospects for future research: Even though empirical research on digital astroturfing is difficult, it is neither impossible nor futile.

Keywords

political communication, Internet, public opinion, astroturfing, bots

1 Introduction

Not long after the advent of the World Wide Web, there were high hopes that this new form of access to the Internet might prove a boon for democracy. Thanks to the World Wide Web, citizens could gain access to politically relevant information far easier than ever before. Even more than that: The World Wide Web seemed to carry the potential to make the elusive ideal of deliberative democracy a reality (or, at least, to make instances of deliberative democratic participation easier) (Buchstein, 1997; Dahlberg, 2001; Rheingold, 1993). The positive force of the Internet, the hope went, would make democratic participation easier than ever and thus it would promote democratic values, even in non-democratic countries.

The rapid progress both of Internet penetration and of Internet-related technologies has created circumstances in which the democratic online revolution as envisioned in the World Wide Web's early days could, in principle, take place. Access to the Internet is increasingly common even in relatively poor countries, and both the means for accessing the Internet as well as the ways in which users can access political information and engage in political discourse have proliferated.

Of course, the World Wide Web has not produced the democratic utopia that some might initially have hoped for. But Internet has certainly had an impact on politics and, more specifically, on political discourse. Citizens do indeed seek political information online and they do engage in new forms of digital political discourse. Unfortunately, the relative ease of engaging in political discourse on the Internet makes it also relatively easy to undermine, or at least negatively affect, that very discourse. One perfidious way in which this can happen: Political actors can create online activity that seems like authentic

https://doi.org/10.24434/j.scoms.2018.01.005

activity by regular citizens, when it is, in reality, anything but.

Online users who are not what they pretend to in order to ruse real people are the opposite of what is hoped for in the ideal of deliberative democracy. Fake online users are not interested in participating in rational political discourse, but rather in deliberately and actively skewing the discourse according to the goals they are pursuing. This is not an entirely new tool. So-called astroturfing, fake grassroots campaigns and organizations that are created and backed by political actors or businesses, have a long tradition (Lyon & Maxwell, 2004; Walker & Rea, 2014). Astroturfing on the Internet is more problematic than the traditional kind: Digital astroturfing is cheaper, has a greater scope, and is potentially much more effective than regular astroturfing.

1.1 A new era of political astroturfing

In the wake of the 2016 U.S. presidential election, the National Intelligence Council at the Office of the Director of National Intelligence released a declassified version of an assessment of Russian state-sponsored activities during the election (National Intelligence Council, 2017). In that report, the National Intelligence Council describes the Russian interference in the 2016 election as a continuation of covert influence campaigns that Russia and the Soviet Union were conducting in the past. However, the report also mentions the activities of the St. Petersburg-based "Internet Research Agency" as part of the Russian influence campaign. The Internet Research Agency is a state-sponsored online astroturfing organization (Chen, 2015; MacFarquhar, 2018) specialized in creating and maintaining sock puppets, fake online personae that are mimicking regular users (Bu, Xia, & Wang, 2013). The astroturfing campaign conducted by the Internet Research Agency was great in scope. In early 2018, Twitter notified over 1.4 million of its US-based users that they had interacted with sock puppet accounts likely affiliated with the Internet Research Agency (Twitter PublicPolicy, 2018). But the campaign was also sophisticated, as evidenced by the persuasive power of some of sock puppets, such as the influential, yet fictitious "Jenna Abrams" (Colins & Cox, 2017). The astroturfing campaign of 2016 was so pervasive that it was proposed that the US create a military unit in order to defend from such campaigns (Hart & Klink, 2017).

Digital astroturfing is not limited to U.S. elections. Instances of digital astroturfing have been documented in many countries (Wooley, 2016). Astroturfing efforts can be conducted by outside political actors, such as in the 2017 French election (Ferrara, 2017) or in the 2016 Brexit vote in the United Kingdom (Bastos and Mercea, 2017 ; Llewellyn et al., 2018), or they can be campaigns of political actors within their own country, such as the systematic digital astroturfing strategy of China (King et al., 2017). Digital astroturfing is also not limited to manually curated sock puppets. For example, many digital astroturfing efforts are conducted with automated bots, whereby computer software impersonates humans by acting and reacting in as "natural" ways as possible (Bessi & Ferrara, 2016; Shao et al., 2017; Wooley, 2016).

1.2 The goal of this paper

The phenomenon of digital astroturfing is of obvious importance. Digital astroturfing represents an inherently deceptive political activity that is detrimental to the democratic discourse and to democratic decision-making. In view of this problem, the goal of this present paper is threefold.

First, we want to define digital astroturfing in a generalized, universally applicable manner. Digital astroturfing is a complex political activity. Second, we propose a typology of different forms of digital astroturfing. That typology is based on the type of actor that is engaging in the astroturfing effort; the persuasion target that is to be affected by the astroturfing effort, and the goal that is pursued by the astroturfing effort. Third, we discuss possible countermeasures to either prevent digital astroturfing or to lessen its impact.

2 A generalized definition of digital astroturfing in politics

In the previous section, we have provisionally introduced digital astroturfing as a type of political online activity that pretends to be the activity of individual, regular online users, when it is in fact activity created by political actors. Such a provisional definition already describes the nature of digital astroturfing to some degree, but it is derived from an intuitive inductive process of describing what different cases of digital astroturfing have in common. A more precise deductive definition is needed in order to describe digital astroturfing in a generalized way, independent from specific singular occurrences of digital astroturfing. To that end, we propose the following definition:

Digital astroturfing is a form of manufactured, deceptive and strategic top-down activity on the Internet initiated by political actors that mimics bottom-up activity by autonomous individuals.

In this definition, digital astroturfing as a form of political activity consists of five necessary conditions: It takes place on the Internet, it is initiated by political actors, it is manufactured, it is deceptive and it is strategic. Digital astroturfing is manufactured in that it is not an honest expression of autonomous individual opinions, but rather activity created to mimic honest expression of autonomous individual opinions. It is deceptive, because the goal of digital astroturfing is to trick the target, usually the public at large (including individuals and small groups), into believing that the manufactured activity is real. It is strategic, because the political actors who engage in digital astroturfing pursue certain goals in doing so. This definition is relatively simple, but in simplicity, we believe, lies analytical clarity. So far, only few authors have attempted to define digital astroturfing. The most relevant such attempt is a recent conceptual study by Zhang, Carpenter, and Ko (2013). In that study, the authors define online astroturfing (the authors use the term online rather than digital astroturfing) as follows: "Therefore, we define online astroturfing as *the dissemination of deceptive opinions by imposters posing as autonomous individuals on the Internet with the intention of promoting a specific agenda"* (p. 3).

This definition is similar to the one we introduce in three respects. First, we have adopted the authors' very apt notion of "autonomous individuals" into our own definition because it describes very precisely what kind of deception is taking place. We expand the descriptive notion of autonomous individuals by contrasting bottom-up and top-down activity; autonomous individuals engage in bottom-up activity, and digital astroturfing is top-down activity pretending to be bottom-up activity. Second, and this is nontrivial, the authors define digital astroturfing as an activity that is taking place on the Internet. And third, they describe digital astroturfing as a deceptive activity. We agree with those three aspects as necessary conditions, but the definition lacks accuracy and generalizability. The definition is inaccurate because it implies that the relevant actors are the imposters posing as autonomous individuals. However, the truly relevant actors are not the imposters themselves, but rather the political actors initiating the astroturfing efforts. The imposters, such as, for example, the employees of the Internet Research Institute in St. Petersburg, are simply performing tasks on behalf of political actors; they are the means to an end. It is possible that in some cases, the actors who initiate the digital astroturfing in pursuit of some goals will also actively carry out the digital astroturfing effort themselves. For example, a politician might himself create a number of fake Twitter accounts with which he would then follow his own, real account in order to feign popularity of his account. But even in such a scenario, the distinction between who initiates the digital astroturfing effort is clearly separate from how the digital astroturfing effort is actually performed. The origin of digital astroturfing lies with the political actors who pursue certain goals by engaging digital astroturfing, and the means of actually carrying out the digital astroturfing effort can be, and in many instances probably are, separate from those political actors.

Furthermore, the above definition by Zhang et al. (2013) is not fully generalizable because it does not describe a general principle, but rather one empirical scenario of digital astroturfing: Cases in which individuals create fake online personæ, the so-called sock puppets, in order to communicate certain points of view. While such a digital astroturfing scenario might be a very important one, it is not the only possible one as we will later point out in our typology.

2.1 The difference between regular astroturfing and digital astroturfing

Ours is not the first study to tackle the phenomenon of digital astroturfing. However, in the previous section, we only discuss one definition of digital astroturfing from one recent study. We are not willingly disregarding other definitions of digital astroturfing, but the status quo of the conceptual work on digital astroturfing is rather limited: Most studies that look at digital astroturfing do not operate with an explicit definition of digital astroturfing, but they tend to extend the concept of regular astroturfing into the online realm instead. For example, Mackie (2009) describes astroturfing on the Internet in the following terms: "Astroturfing is the use of paid agents to create falsely the impression of popular sentiment (the grass roots are fake, thus the term astroturf, which is artificial grass)" (p. 32). Even though the author goes on to argue that and why the Internet is vulnerable to astroturfing, he does not propose a separate, explicit definition of astroturfing that is occurring on the Internet. This implies that astroturfing is astroturfing, no matter where it takes place. Furthermore, Mackie (2009) is only considering paid agents to be indicative of astroturfing. We believe that unpaid agents can be involved in digital astroturfing as well, as in the case of sympathizers who become involved in digital astroturfing because they honestly support a cause; we delve into this issue in the sections below. In a similar vein as Mackie (2009), Boulay (2012) explicitly addresses the potentials of *digital* astroturfing, but she operates with the definition of regular astroturfing: « L'astroturfing est une stratégie de communication dont la source est occultée et qui prétend, à tort, être d'origine citoyenne » (p. 201).

Is a conceptual differentiation between regular and digital astroturfing, as indirectly proposed by our proposal of an explicit definition of digital astroturfing, reasonable given the fact that several authors analyse digital astroturfing, but only apply a definition of regular astroturfing? We believe it is, because digital astroturfing conceptually clearly differs from regular astroturfing. In our definition of digital astroturfing, we identify five necessary conditions: The activities take place on the Internet, they are initiated by political actors, and they are manufactured, deceptive and strategic. If digital astroturfing is to be understood simply as the extension of astroturfing onto the Internet, then four of these five criteria have to apply to regular astroturfing; obviously, regular astroturfing does not have to fulfil the condition of taking place on the Internet. However, regular astroturfing is not adequately described by the remaining four necessary conditions. In regular astroturfing, some political actors are present, and those actors act strategically in that they pursue goals. But those actors are not necessarily always initiating the activities, they are not necessarily directly manufacturing the activities, and they are not necessarily deceitful about their involvement in the activities. Vergani (2014) makes this point clear by introducing the concept of *grassroots orchestra,* a situation where genuine grassroots activity takes place, but over time, political actors get involved in that genuine activity and they influence, if not usurp the grassroots activity in order to steer it in a direction beneficial to themselves. A recent prominent example of a grassroots orchestra is the Tea Party movement in the United States: It seems that the Tea Party was an amalgamation of genuine grassroots activity with coordinating and mobilizing influence by political actors (Formisano, 2012).

The distinction between regular and digital astroturfing, in summary, lies in the fact that digital astroturfing is a clearly demarcated dichotomous phenomenon, while regular astroturfing is not; in regular astroturfing, varying degrees of genuine grassroots components can be present. Of course, this conceptual difference between digital astroturfing and regular astroturfing does not preclude the possibility that there are instances of astroturf campaigns where regular and digital astroturfing are combined in order to achieve as great an impact as possible (Walker, 2014).

2.2 A typology of digital astroturfing

With our generalized definition of digital astroturfing, it is possible to reflect upon what digital astroturfing can empirically look like. Because of the clandestine nature of digital astroturfing, it could be very misleading to base a typology of digital astroturfing solely on the exposed cases of digital astroturfing, since there is no way of knowing whether those exposed cases, such as the briefly mentioned ones in the introduction, cover all possible types of digital astroturfing. For that reason, we opt for a theoretico-deductive typology of digital astroturfing: It is not possible to inductively gather information on all instances of digital astroturfing, but it is possible to define a set of dimensions that will plausibly exhaust the possible types of digital astroturfing.

We propose three dimensions that encompass the different possible types of digital astroturfing: The political actors who engage in digital astroturfing, the target of the digital astroturfing effort, and the goal of digital astroturfing.

The first dimension of digital astroturfing types are the political actors who engage in digital astroturfing. It is reasonable use the category of the political actors that engage in digital astroturfing as a starting point for a typology of digital astroturfing – after all, political actors are the conditio sine qua non of digital astroturfing; political actors are the initiating force behind every digital astroturfing effort. We distinguish between five categories of political actors: governments, political parties, interest groups, individual politicians, and individual citizens. The first two actor categories, governments and political parties, are self-explanatory. The third actor category, interest groups, perhaps less so. We use the term interest group to mean a group that pursues political goals, but is neither part of the government nor a political party. This can include loose groups of like-minded people, nonprofit organizations, businesses, and so forth. Interest groups are thus groups that pursue political interests, and to that end, they interact with the political system. Sometimes, interest group activity will be in the form of discrete, non-public direct interaction with members of government or parliament; i. e., in the form of lobbying. But in other situations, interest groups are advocating their position publicly. The nature of interest groups can be very varied; typical examples of interest groups in Western countries are corporations as employer interest groups and unions as employee interest groups (Walker, 1983). In the context of interest groups as political actors, it is important to keep in mind that digital astroturfing is concerned with political goals. There is a prominent phenomenon of non-political, strictly commercial quasi-astroturfing that some businesses engage in: The practice of faking online consumer reviews to make one's products or services appear more popular than they actually are or to make competitors' products or services appear less popular than they actually are (Yoo & Gretzel, 2009). This type of deceptive online activity by commercial interest groups is not digital astroturfing, because the nature of the activity that the actors pursue is not political. If, however, a business is pursuing political goals and is engaging in digital astroturfing to that end, then we subsume that business in the interest group category of initiating political actors. The fourth category of political actors are individual politicians. Individual politicians are, for the most part, members of groups such as political parties, but there are situations in which individual politicians act on their own in terms of digital astroturfing. We mentioned one such scenario in the introduction: A politician who, as part of his election campaign, paid for Twitter followers.

The second dimension of digital astroturfing types is the persuasion target of digital astroturfing. We differentiate among two general persuasion targets, the public and specific political actors. It is quite obvious to imagine the general public as one target of digital astroturfing, since most known cases of digital astroturfing are cases of public digital astroturfing activities. Of course, political actors can also perceive public digital astroturfing and be affected by it. But another category of digital astroturfing is possible, one that is specifically aimed at political actors and is not public.

The third dimension of the typology of digital astroturfing are the goals the political actors pursue by engaging in digital astroturfing. There are, of course, a myriad specific goals that specific political actors can pursue in specific political contexts. The idea of this third dimension of the digital astroturfing typology is not to describe every possible situational configuration, but rather to meaningfully divide the spectrum of the goals of digital astroturfing into a small number of categories. We identify two such categories of goals: support for or opposition to policy, and support for or opposition to political actors. We propose only two categories, since the goal of digital astroturfing is to communicate a valence either towards policy or towards political actors, and being positive or negative is simply the property of the valence. As for the objects of the valence, political actors who engage in digital astroturfing will either want to influence public opinion or specific political actors with regard to certain policy issues, either to support their own position on those issues, or to oppose the position of other political actors. The other broad category of goals is support for or opposition to political actors. When political actors pursue this goal, they will either do so in order to feign public support for themselves, or to make other political actors appear unpopular.

In public discussions of digital astroturfing, instances of digital astroturfing are sometimes described as *trolling*. To equate digital astroturfing with trolling in such a manner is a grave conceptual error. We have defined digital astroturfing as manufactured, deceptive and strategic activity on the Internet initiated by political actors. Of these five necessary conditions, trolling only satisfies two: Trolling takes place on the Internet, and there is a strategic component to it. However, trolling is not initiated on behalf of political actors, it is not manufactured and it is only somewhat deceptive.

Trolling is best understood as malicious, disruptive or disinhibited (Suler, 2004) online behavior by individuals who engage in the activity of trolling of their own individual volition (Hardaker, 2010). Trolling often also appears as a social, coordinated activity (MacKinnon & Zuckerman, 2012). Still, conceptually, digital astroturfing is completely separate from trolling, and mixing these concepts should be avoided, not least because talking about digital astroturfing in terms of trolling trivializes digital astroturfing: Digital astroturfing is a deceptive effort aimed at the public and launched by political actors. However, there is one potential connection between trolling and digital astroturfing: Cases of digital astroturfing in which the deceptive activity takes the form of trolling. In such cases, the trolling that is taking place is not authentic, but manufactured and strategically used by political actors. We believe this conceptual distinction is important. Not every activity that is widely described as trolling is in fact trolling.

However, it is possible that political actors engage in digital astroturfing and the honest bottom-up activity they mimic is trolling. This behavior can still be described in terms of our typology, because a deceptive and strategic use of trolling is meant to disrupt some public debate and expression of opinion that a political actor objects to, and those public debates and expressions of opinion will, naturally, refer to some policy or to some political actors. Fake trolling thus has the goal of indirectly stymieing expressions of support for or criticism against policy or for political actors. For example, the Chinese government regularly tries to distract the

Table 1: Typology of digital astroturfing

Initiating political actor	Persuasion target	Goal: Support for or opposition to:	
		Policy	Political actors
Government	The Public	1	2
Political party		3	4
Individual politician		5	6
Interest group		7	8
Government	Political actors	9	10
Political party		11	12
Individual politician		13	14
Interest group		15	16

Note. Each number represents one digital astroturfing scenario as a combination of three dimensions. Reading example: The digital astroturfing type with the number 7 describes a digital astroturfing scenario in which an interest group has the goal of influencing the public opinion on some policy issue.

public and change the subject of critical discussions online with fabricated user comments (King et al., 2017).

Combining the three dimensions of the typology leads to the sixteen different types of digital astroturfing, as reported in Table 1, provide a useful framework for analyzing individual cases of digital astroturfing as well as for guiding expectations about digital astroturfing efforts in general. Digital astroturfing can occur, as we argue in the introduction, in very different political contexts, and as a consequence, it can be challenging to analyze separate cases. With our typology, we provide, in essence, a useful heuristic – a way to think about digital astroturfing.

2.3 Digital astroturfing repertoires

Our typology consists of sixteen digital astroturfing scenarios that describe which actors pursue what kind of goal with their digital astroturfing efforts. However, the different scenarios do not automatically imply what specific measures the political actors take in order to carry out their digital astroturfing efforts. These specific efforts consist of three elements: The specific digital astroturfing tools used, the specific venues where these tools are applied, and the specific actions that are taken with those tools in those venues.

In social movement research, the concept of *protest repertoires* is used to describe which tools social movements use in which contexts (Della Porta, 2013; Tarrow, 2011). It is useful to use an analogous concept for digital astroturfing: The concept of *digital astroturfing repertoires*. Digital astroturfing repertoires cannot be defined universally, because the tools and venues available for digital astroturfing are very much bound by time and space and are likely to change – just like the protest repertoires of social movements are bound by time and space and likely to change over time (Biggs, 2013). Also, describing digital astroturfing repertoires is ultimately an inductive task of continued observation of digital astroturfing cases. Even though we cannot specify a definitive and final list of context specific digital astroturfing repertoires, there are some typical repertoires, we believe, that encompass a large portion of contemporary digital astroturfing repertoires.

Digital astroturfing repertoires can be thought of, as mentioned above, as combinations of tools, venues and actions. Some typical digital astroturfing tools are sock puppets, click farms, and sympathizers. *Sock puppets* are, as we describe in the introduction, fake online personae (Bu et al., 2013) that can be used for a variety of purposes. Click farms are a relatively recent variant of sock puppets tied to the rise of social media. *Click farms* provide services that, essentially, boil down to faux social media activity, such as fake followers on Twitter or fake likes on Facebook (Clark, 2015). Both sock puppets and click farms can contain various degrees of automation through *bots,* programs that operate automatically on behalf of agents (Geer, 2005). For example, fake user profiles used in click farms are usually created manually, but the accounts are subsequently operated automatically. A lot of the current research in the area of digital astroturfing is focused on bots. Bots are certainly important, not least quantitatively, since creating and deploying bots is fairly easy and cheap. However, bots are still, for the most part, not able to persuasively mimic humans (Stieglitz et al., 2017), limiting their potential persuasive power. Bots as automated sock puppets or as automated click farms are certainly important, but manually curated and operated sock puppets and click farm profiles still matter. The idea of *sympathizers* as digital astroturfing tools might, prima facie, seem a misclassification. After all, sympathizers are real people who honestly hold the opinions they put forward. The question whether sympathizers as digital astroturfing tools honestly believe in what they do is conceptually irrelevant. Political actors are still orchestrating, i. e., manufacturing some activity that is supposed to look like non-manufactured, spontaneous activity; that activity is deceptive (the target of the astroturfing attempt has no knowledge of its manufactured origins); and the activity is strategic from the point of view of the political actor. Whether sympathizers as the tools of digital astroturfing sympathize with the political actors on whose behalf they act has no conceptual impact on the nature of the activity at hand. Howard (2006) describes such a case in which sympathizers were mobilized in an orchestrated attempt without even knowing that they were instrumentalized by a political actor. Finally, paid supporters are similar to sympathizers: They both engage in activities on behalf of political actors without openly declaring so, and they both do so with their true identities rather than with sock puppets. However, paid supporters are not (only) acting out of personal conviction, but out of pecuniary interests – they are being paid for their activities.

Some typical venues of digital astroturfing are social media, websites, comment sections (predominantly those of news websites), and emails. We understand *social media* as platforms where users, be they individuals, groups or organizations, can connect with other users (Boyd & Ellison, 2007) and where the users themselves are the primary creators of content (Kaplan & Haenlein, 2010). Typical examples of social media are services such as Facebook, Twitter, Instagram, Snapchat, and so forth. We also consider online petitions to be a form of social media; after all, when signing online petitions, users connect with each other and by doing so create a certain type of content. *Websites* are separate from social media in that the users who curate websites do not simply create profiles on social media sites, but rather create separately hosted, independent content. *Comment sections* are the sections on websites that allow website visitors to leave written comments below content on the website, and sometimes, to like or dislike existing comments. In the context of digital astroturfing, comment sections of news websites are potentially one of the most relevant venues. In principle, comment sections could be thought of as just another form of social media, but we treat them as a separate venue. Users are the primary creators of content on social media, but not in comment sections of news websites – in comment sections, users only react to content created by journalists. The final venue of digital astroturfing are *direct messages.* Describing direct messages (such as emails, online contact forms, private messages on social media, and so forth) as venues is perhaps somewhat confusing, because direct messages aren't public places on the Internet. But what we mean by venue is not a public location, but rather the place where the digital astroturfing tools are applied. However, direct messages are a somewhat special case of digital astroturfing, because direct messages can be targeted at specific political actors, without the general public taking notice.

Finally, there are only two digital astroturfing *actions:* Actively creating content, or passively signaling (dis-)approval. The difference between active content creation and passive signaling of (dis-)approval can be exemplified with comment sections on websites. Writing a comment constitutes an active creation of content, whereas liking or disliking is merely a passive signaling of approval or disapproval. The different tools, the different venues and the different actions of digital astroturfing are summarized in Table 2.

A digital astroturfing repertoire may consist of any combination of tools, venues and actions from Table 2. For example, the combination of sock puppets with comment sections (of news websites) and creating content is the digital astroturfing repertoire used by the Russian Internet Research Agency described in the introduction. The combination of click farms with social media and the signaling of approval is the digital astroturfing repertoire used by the Swiss politician mentioned in the introduction who bought fake twitter followers. A repertoire consisting of sympathizers combined with direct messages and creating content is a digital astroturfing repertoire sometimes applied by political candidates running for office who encourage sympathizers to send pre-written emails to friends and to newspaper editors (Klotz, 2007).

Table 2: Tools, venues and actions of digital astroturfing repertoires

Tools	Venues	Actions
sock puppets	social media	creating content
click farms	websites	signalling (dis-) approval
sympathizers	comment sections	
paid supporters	Direct messages	

3 Countermeasures

Digital astroturfing is a clandestine activity, and as such, it is difficult to devise defense mechanisms against it. While there is probably no way to completely prevent digital astroturfing from occurring, there are ways to limit the probability of digital astroturfing occurring or, at least, of digital astroturfing being effective in some venues. More specifically, we identify two broad strategies of countermeasures: Restrictive countermeasures and incentivizing countermeasures. Restrictive countermeasures are measures that limit the spectrum of possible online activities, and this limitation of online activities in general also limits digital astroturfing activities. Incentivizing countermeasures, in contrast, are measures that are conducive to honest forms of online activity. Incentivizing countermeasures do not necessarily prevent digital astroturfing, but by rewarding more openly honest activity, incentivizing countermeasures make digital astroturfing less prominent and thus, potentially, less impactful.

Before we delve into both types of countermeasures in more detail, it is in order to justify why exactly we propose countermeasures in the first place – the fact that we actively discuss countermeasures implies that we regard digital astroturfing to be a problem.

There are two major reasons why digital astroturfing is an important and problematic phenomenon, a normative one and an empirical one. Normatively, digital astroturfing is challenging because the act of engaging in digital astroturfing, essentially, amounts to subterfuge, and subterfuge is nothing more than a form of lying (Grover, 1993). Empirically, digital astroturfing almost certainly has some real-world impact. While it is, for obvious reasons, very difficult to quantify the effects of digital astroturfing on public opinion or on political actors, it is not far-fetched to assume that some effects are very likely to exist. If, for example, a user comment in the comment section of a news website is written by a sock puppet, and some individual reads that comment without realizing that it was written by a sock puppet, then the minimal effect of digital astroturfing has happened – someone was deceived by the digital astroturfing effort. The probability of such a minimal effect of digital astroturfing is very high; if the amount of people who get deceived by digital astroturfing efforts is greater than zero, then the minimal effect exists, and consequently, digital astroturfing has an empirical impact. What is not clear, however, is the full extent of the empirical impacts of digital astroturfing. For example, there is some evidence that user-generated comments on news articles can, in general, affect the perception of the content of those news articles (Anderson, Brossard, Scheufele, Xenos, & Ladwig, 2014; Lee, 2012). It is entirely possible that astroturfed comments can produce similar effects.

3.1 Restrictive countermeasures, or: Limiting online activity

The most sweeping restrictive countermeasure against digital astroturfing is be to outlaw it. However, an outright criminalization of digital astroturfing is, at the very least, hardly enforceable (digital astroturfing is, after all, clandestine). Also, digital astroturfing activities that do not use sock puppets but individuals who engage in digital astroturfing activities with their real identities, would present a difficult legal situation: While the activity is clearly digital astroturfing (a political actor manufactures deceptive activity in order pursue his own goals), the individuals who represent the digital astroturfing tools can easily claim to honestly have the opinions that they express through their activities. Outlawing digital astroturfing thus seems very unrealistic. Furthermore, such action would potentially harm freedom of speech and thus, from a consequentialist point of view, do more harm than good (e. g., Deibert et al., 2008).

More realistic means of restrictive countermeasures are countermeasures applied at the different digital astroturfing venues as summarized in Table 2. But not all of the venues are suited for restrictive measures. Specifically, digital astroturfing can neither be meaningfully restricted when it comes to creating websites (it is trivially easy to create a website anonymously) nor when it comes to sending direct messages, such as emails, to political actors. But for the other two venues, restrictive countermeasures can be implemented.

The first of those venues is social media. One potential restrictive measure for social media is the detection of inauthentic user accounts. There already exists considerable research in the area of automated detection of fake social media accounts, usually in the context of spam accounts that post links to products and services or to other unrelated online content (Mukherjee, Liu, & Glance, 2012; Zheng, Zeng, Chen, Yu, & Rong, 2015). While there is some research that extends the question of automated detection of fake profiles to digital astroturfing (Chen et al., 2011), that research is still scarce. Another possible restrictive countermeasure for social media platforms is the implementation of a mandatory real-name policy for users. Implementing a mandatory real-name policy might sound drastic, but it has been the official policy of one of the largest social media platforms, Facebook, for years (Wauters, Donoso, & Lievens, 2014; Giles, 2011). However, even though using one's real name is the default policy at Facebook, not every user has to or is able to provide credentials for their identity. Facebook is detecting potential pseudonyms algorithmically and, apparently, selectively, which has led to instances of users accounts being blocked despite being honest expressions of real identities (Dijck, 2013; Lingel & Golub, 2015). Mandatory real-name policies for social media also exist beyond Western social media platforms, notably in China. A prominent step towards banning anonymity on Chinese social media has been taken in 2011, when a mandatory real-name policy for so-called microblogging sites that are headquartered in Beijing, such as Sina Weibo, had been introduced. Under that policy, users have to register with their real identities, but are allowed to use nicknames on their public profiles. However, the policy has not yet been thoroughly enforced (Jiang, 2016).

The second digital astroturfing venue where restrictive countermeasures can be implemented with a high probability of the desired impact are comment sections on news websites. By and large, comment sections of news websites are already governed by some degree of restrictive measures in that most news websites have some moderation policies in place (Canter, 2013; Hille & Bakker, 2014). Comment moderation policies can be made more restrictive in order to prevent astroturfed comments. One restriction measure is to ban anonymous comments and enforce a real-name policy, similar to the social media examples cited above. From a logical point of view, banning anonymous comments will always prevent sock puppets from commenting, which means that this measure is definitively effective, but only if it is systematically enforced by means of an explicit verification process.

Enforcing a real-name policy is not a viable restrictive countermeasure against all digital astroturfing tools listed in Table 2: Paid supporters and sympathizers will still be able to be active on social media and to comment on news websites, since they do so with their real identity.

Another possible restrictive measure is not aimed at the tools used for digital astroturfing, but at the specific actions of digital astroturfing: Disabling the possibility of signaling approval or disapproval for existing comments through liking or disliking them. From a logical point of view, disabling the possibility of liking or disliking comments will always prevent this form of astroturfing. A third and most far-reaching restrictive measure with regard to comments is to completely disable comments. A number of prominent news websites have opted for disabling their comment section in recent years (Finley, 2015), but not primarily because of perceived inauthentic commenters. Disabling comments will, from a logical point of view, always prevent all instances of astroturfed comments and, as a logical consequence, it will always prevent all instances of astroturfed "liking" or "disliking" of existing comments as well. This means that completely disabling comments is the most effective restrictive measure against astroturfed comments, since it always prevents 100% of astroturfing instances.

In summary, there are ways to implement restrictive countermeasures against digital astroturfing which are very likely to lead to the desired outcome, which is reducing or completely preventing instances of digital astroturfing. However, implementing such restrictive measures can also have some downsides, and we believe that the benefit of reducing digital astroturfing by restrictive measures does not necessarily outweigh the downsides. When it comes to the automatic detection of unauthentic user profiles on social media platforms, the potential downside are primarily false positives: When automatic means of detecting and deleting inauthentic user accounts are implemented, there is a chance that a portion of the accounts thus detected will actually be authentic ones. However, this downside of false positives can be actively controlled and dialed down. How many false positives there are is to a substantial degree a trade off with the desired sensitivity of the automatic detection algorithms. If a maximum sensitivity is desired, meaning that a perfect rate of true positives is the goal, then the false positive rate will, probably, be at its highest. But if a lower sensitivity is accepted, meaning that not all true positives will be detected, then the false positive rate can be lowered.

The downsides of restrictive countermeasures in comments sections on news websites are less open to adjustment than the downsides of automatic detection of inauthentic user accounts on social media. The first such measure we discuss above is enforcement of a real-name policy in the comment sections. Perhaps it is not self-evident why non-anonymity should have downsides at all – if one has an honest opinion to publicly share, why not do so using their true identity? The problem with that notion is that in many non-democratic countries, voicing one's opinion publicly can have dire consequences. A prominent example of such a country is China, where online users criticize the government under the guise of anonymity because such criticism is essentially illegal (Qiang, 2011). Limiting the possibilities of anonymous online commenting in such political contexts will either lead to less expression of opinion, or to an increase in sanctions because those who speak out are easier to be tracked down. Another downside of doing away with anonymous commenting is commercial in nature. Removing anonymous commenting reduces the total comment volume (Fredheim, Moore, & Naughton, 2015), but only a small part of those anonymous comments is likely to be astroturfed. This means that websites will lose a lot of user-generated content and, consequently, website traffic. This can ultimately lead to loss in revenue, which is a relevant factor for commercial operators of social media and of news websites.

The strongest measure with regard to comments on news websites that we discuss above is the complete disabling of comments. Completely removing the possibility for users to comment is a measure, we believe, that represents a step backwards in terms of the development of the Internet, because it is a measure that reduces the opportunities for honest citizens to participate in public discourse. Of course, the Internet is a big place, and any one user can, for example, easily create a website to express whatever opinions he or she may hold. Furthermore, news organizations obviously have no obligation to host user-generated content (Jönsson & Örnebring, 2011) of any sort on their online platforms. But the possibility for citizens to directly react to news content by commenting on it and by engaging with other users who comment as well does seem like an added discursive value which is lost by disabling comments altogether.

Restrictive countermeasures against digital astroturfing are possible, but they carry with them some degree of downsides. Because of those downsides, it is worth exploring alternative strategies of countermeasures: Such countermeasures that do not restrict, but instead incentivize.

3.2 Incentivizing countermeasures, or: Rewarding honesty

The general idea behind incentivizing countermeasures against digital astroturfing is to provide incentives that motivate users to behave in ways that are more likely to be honest. Incentivizing honest behavior does not prevent digital astroturfing from occurring, but by rewarding honest behavior, honest online activity can become publicly discernible from online activity for which there is no information about its authenticity.

One goal of such incentives is to encourage user participation under one's true identity, without enforcing it. In the case of social media, encouraging users to use their true identity consists of two steps. First, there needs to be an actual possibility for users to verify their identity on a given social media platform. Second, after a user has verified his or her identity, the information of the authenticity of the social media profile in question needs to be publicly visible. Some prominent social media platforms, such as Twitter and Facebook, have such a verification program in place, and verified accounts are awarded with a publicly visible badge. However, both Twitter and Facebook do not currently implement their verification scheme widely for their whole user base, but only to a relatively small portion of users who are a "public figure, media company or brand" (Facebook, 2015). We believe that users can be incentivized to opt into a verification process by nudging them towards it. Users can be presented with advantages of a verified account, such as the pro-social norm of transparency and authenticity, but also with the rather playful element of receiving a badge that is proof of their verified status. However, such nudging strategies are, to a certain degree, paternalistic in nature (Thaler & Sunstein, 2009). We think that the paternalistic element of verification options should be minimized by explicitly presenting verification as a choice that brings with it certain benefits.

User verification is also possible for the venue of comment sections. Besides identity verification, another type of verification is possible for comment sections: Incentives that allow for anonymous commenting, but reward a history of good and consistent commenting behavior. One such scheme is currently applied by the New York Times, where some commenters are algorithmically selected for a publicly visible verification badge based on their commenting history (Sullivan, 2014). Commenters become eligible for verification if they have not run afoul of comment moderators in the past. The main incentive under that scheme is that verified commenters are able to immediately get their comments published, without prior moderation.

In comparison with restrictive countermeasures, incentivizing ones have the great advantage that they do not limit the spectrum of online activities, but rather expand it by offering rewards for behavior that is more likely to be honest. However, incentivizing countermeasures against digital astroturfing are not without downsides. One major downside is the temporal component of incentivization: It takes time, first, to implement incentives, and second, for a critical part of a given user base to react to incentives. For example, if Twitter and Facebook decided to offer verification to all of their users, that would pose a non-trivial technical challenge, and the widespread adoption of verification would probably not be immediate but rather follow adoption rates and patterns similar those of other technological innovations (Rogers, 2003).

4 Conclusion: Is research on digital astroturfing feasible?

Laying out a conceptual map of digital astroturfing is only a means to an end – a map is meant to be used to explore. We have described a typology with ideal-types (Weber, 1922). In the real world, no two instances of digital astroturfing will be exactly the same, and they might employ any mix of astroturfing repertoires. Still, the map that we propose, even though preliminary and likely incomplete, should be of some use for maneuvering the terrain of digital astroturfing. But how exactly can digital astroturfing be explored empirically? After all, digital astroturfing is a clandestine activity, and if it is carried out successfully, we do not know that it has taken place. This makes any kind of research inherently challenging. For example, research designs that are routinely used in the study of other forms of political communication can be impossible to implement in the study of digital astroturfing, since very basic facts such as who is doing what in the pursuit of which strategic goals is, by definition, absent in digital astroturfing. Even though it undoubtedly poses unique challenges, the empirical study of digital astroturfing is not futile.

A first step in the study of digital astroturfing is the establishment of a plausible conceptual framework of digital astroturfing. The very goal of our study is to contribute to this first step. In order to conduct empirical research, we first need a sound understanding of how to think about our object of study. In this sense, we do not think that research on digital astroturfing should be exploratory in nature, as is sometimes suggested for research on regular astroturfing (Boulay, 2013).

Some important recent research efforts in the area of digital astroturfing are focusing on so-called "social bots" (Bastos and Mercea, 2017; Ferrara et al., 2016; Woolley, 2016;). Social bots represent one digital astroturfing tool (automated sockpuppets) used within one venue (social media). As such, they most certainly warrant scientific scrutiny. However, it is important to note that social bots probably represent the low hanging fruit of digital astroturfing. Social bots are a relatively crude form of digital astroturfing, and because of their highly automated nature, they are relatively easy to detect. Furthermore, and perhaps just as importantly, the less sophisticated a digital astroturfing effort, the less probable it is that the effort will have the impact intended by the initiating actor behind the digital astroturfing effort. Not only could there be no effect, but there could actually be a negative effect (from the point of view of the initiating actor): When people become aware of persuasion attempts and thus develop persuasion knowledge, they tend not only to not be persuaded, but to actually react negatively (Friestad & Wright, 1994).

There are several empirical research strategies that can be applied to the study of digital astroturfing. A very important one are case studies. To date, most research on cases of digital astroturfing is done by investigative journalists. Such journalistic work is, of course, very valuable, but a more scientifically vigorous analysis of digital astroturfing cases is still necessary, not least because of the need to ground case study research in a sound conceptual framework. An obvious disadvantage of case studies is the fact that the cases that can be analyzed cannot be freely selected – only those instances of digital astroturfing that have been revealed to be digital astroturfing are available for case study analysis. This a priori selection could potentially introduce bias, because observed digital astroturfing is perhaps different from unobserved digital astroturfing. However, this potential bias is not an insurmountable problem. Given a conceptual foundation that consists, among other things, of a typology and a reasonable expectation of digital astroturfing repertoires, the potential selection bias in case study research can be actively addressed. For example, if case study research consistently failed to identify a certain type of digital astroturfing that could indicate that that type of digital astroturfing is very successful at remaining hidden.

Another possible research strategy consists of surveys and interviews of political actors. It might seem naive to suggest to simply ask political actors whether they engage in digital astroturfing. However, it is not at all unheard of for political actors to disclose sensitive information for scientific purposes as long as they are guaranteed anonymity. But talking to political actors about digital astroturfing does not have to produce a direct "admission of guilt" to be useful. For example, the perception of digital astroturfing by political actors is valuable data, because it could indicate how prevalent a phenomenon digital astroturfing is. Talking to people who are potentially involved in or knowledgeable about digital astroturfing does not have to be limited to political actors. Communication professionals in the business of political consulting are also willing to talk about their line of work, given a sufficient layer of anonymity (e.g., Serazio, 2014).

A third promising research strategy is the direct collaboration with venues of digital astroturfing. Two of the venues that are summarized in Table 2 are suitable for this, social media services and news organizations that have comment sections on their websites. Both social media services and news organizations should, in principle, be interested in reducing, or at least in understanding digital astroturfing on their respective platforms. The specific nature of the collaboration can be twofold. First, the combined effort of researchers and operators of social media or of news platforms can be focused on the detection of digital astroturfing. The second possible kind of collaboration is the implementation and monitoring of countermeasures against digital astroturfing. For this kind of collaboration, greater weight should be given to the implementation of incentivizing countermeasures than to restrictive ones. Restrictive countermeasures are mostly a technical affair with potentially big downsides, while incentivizing countermeasures involve not only technical innovation, but observation of and interaction with users as well. Restrictive countermeasures carry the promise of relatively fast short-term solutions, but in order to effectively combat digital astroturfing in the long term, we believe that innovations in the form of incentivizing countermeasures are inevitable.

References

Anderson, A.A., Brossard, D., Scheufele, D.A., Xenos, M.A., & Ladwig, P. (2014). The "Nasty Effect": Online incivility and risk perceptions of emerging technologies. *Journal of Computer-Mediated Communication, 19*(3), 373–387. doi:10.1111/jcc4.12 009.

Apfl, S., & Kleiner, S. (2014). Die Netzflüsterer [The net whisperers]. *DATUM, 2014*(11). Retrieved from https://datum.at/die-netzfluesterer/.

Bastos, M.T., & Mercea, D. (2017). The Brexit botnet and user-generated hyperpartisan news. *Social Science Computer Review.* Advanced online publication, 1–18. doi:10.1177/0894439317734157.

Bessi, A., & Ferrara, E. (2016). Social bots distort the 2016 U.S. Presidential election online discussion. *First Monday, 21*(11). Retrieved from http://firstmonday.org/article/view/7090/5653.

Biggs, M. (2013). How repertoires evolve: The diffusion of suicide protest in the twentieth century. *Mobilization: An International Quarterly, 18*(4), 407–428. doi:10.17813/maiq.18.4.njnu779530x55082.

Boulay, S. (2013). Can methodological requirements be fulfilled when studying concealed or unethical research objects? The case of astroturfing. *ESSACHESS – Journal for Communication Studies, 6*(2/12), 177–187.

Boulay, S. (2012). Quelle(s) considération(s) pour l'éthique dans l'usage des technologies d'information et de communication en relations publiques? Analyse de cas d'astroturfing et réflexion critique. *Revista Internacional de Relaciones Públicas, 2*(4), 201–220.

Boyd, D.M., & Ellison, N.B. (2007). Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication, 13*(1), 210–230. doi:10.1111/j.1083-6101.2007.00393.x.

Bu, Z., Xia, Z., & Wang, J. (2013). A sock puppet detection algorithm on virtual spaces. *Knowledge-Based Systems, 37* (January), 366–377. doi:10.1016/j.knosys.2012.08.016

Buchstein, H. (1997). Bytes that bite: The Internet and deliberative democracy. *Constellations, 4*(2), 248–263. doi:10.1111/1467-8675.00052.

Canter, L. (2013). The misconception of online comment threads. *Journalism Practice, 7*(5), 604–619. doi:10.1080/17512786.2012.740172.

Chen, A. (June 2, 2015). The agency. *The New York Times*. Retrieved from http://www.nytimes.com/2015/06/07/magazine/the-agency.html.

Chen, C., Wu, K., Srinivasan, V., & Zhang, X. (2011). Battling the Internet water army: Detection of hidden paid posters. *arXiv:1111.4297 [cs]*. arXiv: 1111.4297.

Clark, D.B. (April 21, 2015). The bot bubble: How click farms have inflated social media currency. *The New Republic*. Retrieved from https://newrepublic.com/article/121551/bot-bubble-click-farms-have-inflated-social-media-currency.

Dahlberg, L. (2001). The Internet and democratic discourse: Exploring the prospects of online deliberative forums extending the public sphere. *Information, Communication & Society, 4*(4), 615–633. doi:10.1080/13691180110097030.

Deibert, R., Palfrey, J., Rohozinski, R., & Zittrain, J. (Eds.) (2008). *The information revolution and global politics. Access denied: The practice and policy of global Internet filtering*. Cambridge, MA: MIT Press.

Della Porta, D. (2013). Repertoires of contention. In D.A. Snow, D. Della Porta, B. Klandermans, & D. McAdam (Eds.), *The Wiley-Blackwell encyclopedia of social and political movements*. Oxford, UK: Blackwell. doi:10.1002/9780470674871.wbespm178.

DePaulo, B.M., Kashy, D.A., Kirkendol, S.E., Wyer, M.M., & Epstein, J.A. (1996). Lying in everyday life. *Journal of Personality and Social Psychology, 70*(5), 979–995. doi:10.1037/0022-3514.70.5.979.

Dijck, J.v. (2013). "You have one identity": Performing the self on Facebook and LinkedIn. *Media, Culture & Society, 35*(2), 199–215. doi:10.1177/0163443712468605

Erat, S., & Gneezy, U. (2011). White lies. *Management Science, 58*(4), 723–733. doi:10.1287/mnsc.1110.1449.

Facebook (2015). *What is a verified page or profile?* Retrieved from https://www.facebook.com/help/196050490547892.

Ferrara, E. (2017). Disinformation and social bot operations in the run up to the 2017 French presidential election. *First Monday, 22*(8). Retrieved from https://arxiv.org/ftp/arxiv/papers/1707/1707.00086.pdf.

Fielding, N., & Cobain, I. (March 17, 2011). Revealed: US spy operation that manipulates social media. *The Guardian*. Retrieved from http://www.theguardian.com/technology/2011/mar/17/us-spy-operation-social-networks.

Finley, K. (October 8, 2015). A brief history of the end of the comments. *WIRED*. Retrieved from http://www.wired.com/2015/10/brief-history-of-the-demise-of-the-comments-timeline/.

Formisano, R.P. (2012). *The Tea Party: A brief history*. Baltimore, MD: Johns Hopkins University Press.

Fredheim, R., Moore, A., & Naughton, J. (2015). *Anonymity and online commenting: An empirical study* (SSRN Scholarly Paper No.

ID 2591299). Social Science Research Network. Rochester, NY.

Geer, D. (2005). Malicious bots threaten network security. *Computer, 38*(1), 18–20. doi:10.1109/MC.2005.26.

Giles, J. (2011). Zeroing in on the real you. *New Scientist, 212*(2836), 50–53. doi:10.1016/S0262-4079(11)62669-9.

Grover, S.L. (1993). Lying, deceit, and subterfuge: A model of dishonesty in the workplace. *Organization Science, 4*(3), 478–495. doi:10.1287/orsc.4.3.478.

Hardaker, C. (2010). Trolling in asynchronous computer-mediated communication: From user discussions to academic definitions. *Journal of Politeness Research. Language, Behaviour, Culture, 6*(2), 215–242. doi:10.1515/JPLR.2010.011.

Hart, S.W., & Klink, M.C. (2017). 1st Troll Battalion: Influencing military and strategic operations through cyber-personas. In *2017 International Conference on Cyber Conflict (CyCon U.S.)* (pp. 97–104). doi:10.1109/CYCONUS.2017.8167503.

Hille, S. & Bakker, P. (2014). Engaging the social news user. *Journalism Practice, 8*(5), 563–572. doi:10.1080/17512786.2014.899758

Howard, P.N. (2006). *New media campaigns and the managed citizen. Communication, society and politics.* Cambridge: Cambridge University Press.

Jiang, M. (2016). Managing the micro-self: the governmentality of real name registration policy in Chinese microblogosphere. *Information, Communication & Society, 19*(2), 202–220. doi:10.1080/1369118X.2015.1060723.

Jönsson, A.M., & Örnebring, H. (2011). User-generated content and the news. *Journalism Practice, 5*(2), 127–144. doi:10.1080/17512786.2010.501155.

Kaplan, A.M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of social media. *Business Horizons, 53*(1), 59–68. doi:10.1016/j.bushor.2009.09.003.

King, G., Pan, J., & Roberts, M.E. (2017). How the Chinese government fabricates social media posts for strategic Distraction, not engaged argument. *American Political Science Review, 111*(3), 484–501. doi:10.1017/S0003055417000144.

Klotz, R.J. (2007). Internet campaigning for grassroots and astroturf support. *Social Science Computer Review, 25*(1), 3–12. doi:10.1177/0894439306289105.

Lee, E.-J. (2012). That's not the way it is: How user-generated comments on the news affect perceived media bias. *Journal of Computer-Mediated Communication, 18*(1), 32–45. doi:10.1111/j.1083-6101.2012.01597.x.

Lingel, J., & Golub, A. (2015). In face on Facebook: Brooklyn's drag community and sociotechnical practices of online communication. *Journal of Computer-Mediated Communication, 20*(5), 536–553. doi:10.1111/jcc4.12125.

Llewellyn, C., Cram, L., Favero, A., & Hill, R.L. (2018). For whom the bell trolls: Troll behaviour in the Twitter Brexit debate. *ArXiv:1801.08754 [Cs]*. Retrieved from http://arxiv.org/abs/1801.08754.

Lyon, T.P., & Maxwell, J.W. (2004). Astroturf: Interest group lobbying and corporate strategy. *Journal of Economics & Management Strategy, 13*(4), 561–597. doi:10.1111/j.1430-9134.2004.00023.x.

MacFarquhar, N. (February 18, 2018). Inside the Russian troll factory: Zombies and a breakneck pace. *The New York Times.* Retrieved from https://www.nytimes.com/2018/02/18/world/europe/russia-troll-factory.html

Mackie, G. (2009). Astroturfing infotopia. *Theoria: A Journal of Social and Political Theory, 56*(119), 30–56.

MacKinnon, R., & Zuckerman, E. (2012). Don't feed the trolls. *Index on Censorship, 41*(4), 14–24. doi:10.1177/0306422012467413.

McCornack, S.A. & Levine, T.R. (1990). When lies are uncovered: Emotional and relational outcomes of discovered deception. *Communication Monographs, 57*(2), 119–138. doi:10.1080/03637759009376190.

National Intelligence Council. (2017). *Assessing Russian Activities and Intentions in Recent US Elections* (Intelligence Community Assessment). Washington, DC. Retrieved from https://www.dni.gov/files/documents/ICA_2017_01.pdf.

Neuhaus, C. (September 2, 2015). Hans-Peter Portmanns Twitter-Wunder: 11 000 Freunde sollt ihr sein. *Neue Zürcher Zeitung.*

Qiang, X. (2011). The battle for the Chinese Internet. *Journal of Democracy, 22*(2), 47–61. doi:10.1353/jod.2011.0020.

Rheingold, H. (1993). *The virtual community: Homesteading on the electronic frontier.* Cambridge, MA: MIT Press.

Rogers, E. M. (2003). *Diffusion of innovations.* New York: Free Press.

Serazio, M. (2014). The new media designs of political consultants: Campaign production in a fragmented era. *Journal of Communication, 64*(4), 743–763. doi:10.1111/jcom.1 2078.

Shao, C., Ciampaglia, G. L., Varol, O., Flammini, A., & Menczer, F. (2017). The spread of misinformation by social bots. *ArXiv:1707.07592 [Physics].* Retrieved from http://arxiv.org/abs/1707.07592.

Stieglitz, S., Brachten, F., Berthelé, D., Schlaus, M., Venetopoulou, C., & Veutgen, D. (2017). Do social bots (still) act different to humans? – Comparing metrics of social bots with those of humans. In G. Meiselwitz (eds), *Social computing and social media. Human behavior* (pp. 379–395). Cham, Switzerland: Springer. doi:10.1007/978-3-319-58559-8_30.

Suler, J. (2004). The online disinhibition effect. *CyberPsychology & Behavior, 7*(3), 321–326. doi:10.1089/1094931041291295.

Sullivan, M. (April 28, 2014). How “verified commenters” earn their status at The Times, and why. *The New York Times.* Retrieved from http://publiceditor.blogs.nytimes.com/2014/04/28/how-verified-commenters-earn-their-status-at-the-times-and-why/.

Tarrow, S. G. (2011). *Power in movement: Social movements and contentious politics.* Cambridge, MA: Cambridge University Press.

Thaler, R. H., & Sunstein, C. R. (2009). *Nudge: Improving decisions about health, wealth, and happiness.* New York: Penguin.

Twitter PublicPolicy. (2018). *Update on Twitter’s review of the 2016 U.S. election.* Retrieved from https://blog.twitter.com/official/en_us/topics/company/2018/2016-election-update.html.

Vergani, M. (2014). Rethinking grassroots campaigners in the digital media: The “grassroots orchestra” in Italy. *Australian Journal of Political Science, 49*(2), 237–251. doi:10.10 80/10361146.2014.898129

Walker, E. T. (2014). *Grassroots for hire: Public affairs consultants in american democracy.* Cambridge, MA: Cambridge University Press.

Walker, E. T., & Rea, C. M. (2014). The political mobilization of firms and industries. *Annual Review of Sociology, 40*(1), 281–304. doi:10.1146/annurev-soc-071913-043215

Walker, J. L. (1983). The origins and maintenance of interest groups in America. *The American Political Science Review, 77*(2), 390–406. doi:10.2307/1958924.

Wauters, E., Donoso, V., & Lievens, E. (2014). Optimizing transparency for users in social networking sites. *info, 16*(6), 8–23. doi:10.1108/info-06-2014-0026.

Weber, M. (1922). *Wirtschaft und Gesellschaft.* Tübingen, Germany: Mohr.

Woolley, S. C. (2016). Automating power: Social bot interference in global politics. *First Monday, 21*(4).

Yoo, K.-H., & Gretzel, U. (2009). Comparison of deceptive and truthful travel reviews. In D. W. Höpken, D. U. Gretzel, & D. R. Law (Eds.), *Information and Communication Technologies in Tourism 2009* (pp. 37–47). Vienna: Springer. doi:10.1007/978-3-211-93971-0_4

Zhang, J., Carpenter, D., & Ko, M. (2013). Online astroturfing: A theoretical perspective. *AMCIS 2013 Proceedings.*

Zheng, X., Zeng, Z., Chen, Z., Yu, Y., & Rong, C. (2015). Detecting spammers on social networks. *Neurocomputing, 159,* 27–34. doi:10.1016/j.neucom.2015.02.047.