The Politics of Social Media Manipulation

Edited by Richard Rogers and Sabine Niederer

Amsterdam University Press

Amsterdam University Press

The Politics of Social Media Manipulation

Amsterdam University Press

Amsterdam
University
Press

The Politics of Social Media Manipulation

Edited by
Richard Rogers
and Sabine Niederer

Amsterdam University Press

Cover design: Coördesign, Leiden
Typesetting: Crius Group, Hulshout

ISBN 978 94 6372 483 8
e-ISBN 978 90 4855 167 5 (pdf)
DOI 10.5117/9789463724838
NUR 670

Creative Commons License CC BY NC ND (http://creativecommons.org/licenses/by-nc-nd/3.0)

© All authors / Amsterdam University Press B.V., Amsterdam 2020

Some rights reserved. Without limiting the rights under copyright reserved above, any part of this book may be reproduced, stored in or introduced into a retrieval system, or transmitted, in any form or by any means (electronic, mechanical, photocopying, recording or otherwise).

Table of Contents

Amsterdam University Press

List of figures and tables

Figures

Amsterdam University Press

Amsterdam University Press

Amsterdam University Press

Amsterdam University Press

Amsterdam University Press

Amsterdam University Press

Amsterdam University Press

Amsterdam University Press

Amsterdam University Press

Amsterdam University Press

Tables

Amsterdam University Press

Amsterdam University Press

Amsterdam
University
Press

1 The politics of social media manipulation

Richard Rogers and Sabine Niederer

Abstract

This chapter gives an overview of the contemporary scholarship surrounding 'fake news'. It discusses how the term has been deployed politically as a barb against the free press when publishing inconvenient truths since the mid-nineteenth century. It also addresses how such notions have been used in reaction to novel publishing practices, including to the current social media platforms. More generally, the scholarship could be divided into waves, whereby the first related to the definitional issues and the production side, whilst the second has been concerned with its consumption, including the question of persuasion. There is additionally interest in solutions, including the critique of the idea that automation effectively addresses the problems. It concludes with research strategies for the study of the pervasiveness of problematic information across the internet.

Keywords: fake news, junk news, disinformation, clickbait, hyperpartisan, post-truth

Introduction: Influence campaigning in political spaces online and the question of persuasion

In reviewing the scholarship surrounding so-called fake news, one would out of necessity make a distinction between the dominant work on the art of influence campaigning and computational propaganda online and the consequences to date for its consumers, but also the few findings, often journalistic, in the relatively understudied case of the Dutch political space online, both on the web as well as in social media. Much work has

Rogers, Richard, and Sabine Niederer (eds), *The Politics of Social Media Manipulation*. Amsterdam, Amsterdam University Press 2020
DOI: 10.5117/9789463724838_CH01

Amsterdam University Press

been undertaken on the influence of Russian (and Russian-style) influence campaigning in the US, and the presence or absence thereof during elections in Germany, France, Sweden and elsewhere. With respect to the Netherlands, the case studies have been reserved to the early Russian influence campaigning around the downing of the MH17 Malaysian airliner (beginning in 2014) and the suicide bombings in the Brussels airport and metro (2016), discovered through the use of Twitter data sets of Russian trolls, or influence campaigners. Other work has been performed on the existence of home-grown troll networks that are at times 'pro-Russian' but do not seem to have had foreign input.

Crucially, in the studies and journalistic treatments to date it is increasingly remarked that there has been a shift in Russian disinformation campaigning from inflaming conflict with the West to stirring it within the West. It is also argued that disinformation could be said to be 'Russifying', i.e., the borrowing of so-called Russian techniques by domestic actors. The campaigning, whether foreign or domestic, does more than create narratives that divide; it also employs computational means to inflate and amplify them through bot work, fake following, astroturfing, the creation of front groups and other artificial publicity tactics.

It is also argued that more attention ought to be paid to the rise of extreme and divisive media on social media platforms, where the point is often made that great emphasis is being placed on foreign disinformation when by comparison it performs poorly in European news spheres. The growth of 'hyperpartisan' news and commentary also may be viewed as an alternative fact or knowledge infrastructure, contributing to discussions of a post-truth condition and the contention that established institutions are under threat.

It is of equal importance to examine the critique on persuasion, or the extent to which the influence campaigning strategies, artificiality and hyperpartisan sources have discernible impacts on their consumers, especially the voters. They appear to be minimal. Indeed, there is a small, but growing literature critiquing transfer models, also known as hypodermic needle or magic bullet theories which themselves could be considered part and parcel of the fake news hype and fascinations with so-called psyops activities such as in the Cambridge Analytica case.[1] Transfer models do

1 The Cambridge Analytica case or scandal refers to the illegitimate use of over 80 million Facebook users' information to develop micro-targeted advertising (Cadwalladr and Graham-Harrison, 2018). It prompted US Congressional and UK Parliamentary investigations, and also led to Facebook's tightening its data access for academics and public scrutiny more generally (Bruns et al., 2018).

Amsterdam University Press

not take into account the active filtering of media users or phatic sharing, it is argued, whereby one circulates dubious media more to connect with others or for amusement than to pass along substantive information. Such models also would discount hardened attitude, and studies finding that campaigns generally have minimal effects.

As for the measures to be taken, the literature both describes and occasionally questions fact-checking and media literacy efforts because of the assumption that corrected information would assuage knowledge deficits, for attitudes often remain the same. Nonetheless, among the policy recommendations most frequently put forward are bolstering media literacy initiatives, together with flagging questionable content, manually and automatically, for further scrutiny. Social media platforms are facing regulation and are asked to address extreme content and create public archives.

One aspect of the literature review relevant to empirical work concerns the methods employed to demarcate political space online for the subsequent study of the scope and impact of problematic content, junk news and computational propaganda – to use some of the terms for the larger phenomenon under study. Under consideration here are largely mixed (quanti-quali) techniques and digital methods from media studies and data journalism. These provide distinctive political space demarcation strategies for the web as well as social media per platform as well as approaches for cross-platform analysis. They query engines and platforms, measure significant political stories (in terms of engagement) and determine construals of dubiousness through news criticism, categorizing significantly engaged-with stories into genres such as disinformation, conspiracy, clickbait, hyperpartisan and (automated) amplification. While often practiced on a story level, the determination of dubiousness also may be made through source criticism, according to the extent to which publishers' output repeatedly accords with junk news definitions, discussed in the next section. It is also worth studying how signal-based or algorithmic determinations of problematic content comport with qualitative methods that are based on source (provenance) criticism.

Fake news, junk news and computational propaganda

Historically, fake news proclamations could be thought of as linked to particular novel publishing practices both 'when old media were new' but also nowadays through social media platforms (Marvin, 1988; Gitelman, 2006). The term 'canard', meaning unfounded rumour or story, refers to

Amsterdam University Press

the contents printed in the French broadsheets of the eighteenth century; 'scandal sheets' are the British term for the same era of publishing (Darnton, 2010). In the U.S., in particular, 'fake news' as a term recently experienced a revival and travelled internationally, in the numerous senses in which it has been deployed historically: 'news satire, news parody, fabrication, manipulation, advertising, and propaganda' (Tandoc et al., 2018: 137). As directed towards contemporary social media platforms, the charge of fake news and similar terms often has been uttered as a lament after the introduction of new media technologies, where there are concomitant calls for new journalistic standards, as witnessed with the competing tabloids and their sensationalist, yellow journalism in the late 1890s and into World War I as well as the radio and newswire in the 1920s (Opper, 1894; Lippmann, 1922; McQueen, 2018).

With the rise of corporate public relations, the blurring of the distinction between the editorial and the advertisement sent over the wire or into the airwaves prompted the use of the moniker, 'news fakers' (McKernon, 1925; Lazer et al., 2018). Similarly, the contents of the early, unedited web, populated by self-publishers, and later the blogosphere, were often described as 'too fresh to be true', given the speed of news production and the potential for those looking for a scoop to convert unsubstantiated rumour into news (Hall, 2001; Rogers, 2005). More recently, the notion would be routinely deployed by satirical news sources such as *Saturday Night Live!* in the US (Day and Thompson, 2012); in fact, *The Daily Show*, the progressive comedy news program, described itself proudly as a 'fake news program' (Newman, 2010; Harsin, 2018). Parody, it should be recalled, was behind the origination of the most circulated 'fake news' story during the US presidential campaigns of 2015-2016, 'Pope Francis Shocks World, Endorses Donald Trump for President' (Allcott and Gentzkow, 2017). While many definitions concentrate on the falseness of content, they may have the 'trappings of news' through the use of the news apparatus, including the style, formats and containers employed (Laquintano and Vee, 2017; Grinberg et al., 2019). Indeed, narrower definitions take as their point of departure how the sources 'falsely claim to be news organizations,' though they may well look the part (Tucker et al., 2018: 3).

Fake news also has been deployed politically as a barb against the free press when publishing inconvenient truths, or speaking 'truth to power' (Cary, 1955; Darnton, 2017). Since the mid-nineteenth century, labelling the news media generally and disagreeable reporting specifically as the product of *der Lügenpresse* or the lying press is a discrediting ploy or even communication strategy, still practiced today by far-right social movements as Pegida in Germany, chanting at street rallies *Lügenpresse, halt*

Amsterdam University Press

die Fresse (lying press, shut your mouth) (Beiler and Kiesler, 2018). It was the German *Unwort des Jahres* (notorious word of the year) in 2014, in the competition organized by TU Darmstadt. Fake news is also a label, used in highly conservative political circles in the US, for particular news sources, notably CNN, MSNBC, *The New York Times*, and *The Washington Post*; the designation is similar, albeit perhaps more extreme, to past portrayals of the agenda-setting 'elite media' in contrast to conservative upstarts as *Fox News* (Marwick, 2018; Tripodi, 2018; Peck, 2019). In this respect, one could call the current situation just the latest fake news scare, or even moral panic (Brennen, 2017; Morozov, 2017).

When discussing the phenomenon in relation to social media and other online sources, researchers at the computational propaganda project at the Oxford Internet Institute (OII) often offer the umbrella term 'junk news', defined as 'extremist, sensationalist, conspiratorial, masked commentary' (Howard et al., 2017, 1). Other catch-all's include 'problematic information', 'information disorders' and 'false news' (Jack, 2017; Wardle and Derakhshan, 2017). Apart from sensationalist, conspiratorial and masked – features that have been a part of fake news ontologies for centuries – the OII definition emphasizes another element, extremist, which cuts to the heart of contemporary concern for the phenomenon when studied not only as a practice of media and public opinion manipulation but also a trigger for societal unrest.

With respect to the growing anxiety over fake news as harbinger of unrest, one may refer to the distinctions made between a variety of information disorders, as well as the coinage of new terminology that captures excitable, Internet-related media and speech (Wardle, 2018). First, disinformation and misinformation are both false, but the latter is unintentionally so, whilst the former is fashioned for the purposes of intentional disruption and causing harm. A third term, 'mal-information' (a neologism), seemingly borrowed from malware or malicious software categorizations, has been introduced to describe accurate information released for the purposes of harassment such as doxing, or publishing private details (Wardle and Derakhshan, 2017).

These are the tools for the so-called 'weaponization' of social media platforms to foment discord through seeding the news and public opinion with divisive content. Indeed, 'extreme speech' is a term that has been offered as a nuancing of the hate speech discourse as it is applied to online toxicity. It is meant to capture a form of charged language and cultural conflict that stops short of hate, and has emerged with social media, defined as 'vitriolic exchange on Internet-enabled media' (Pohjonen and Udupa, 2017: 1173). Its rise has prompted social media companies as Facebook, Twitter and

Amsterdam University Press

Alphabet (owners of YouTube) to expand their content reviewer pools as well as widen their internal mandates to identify and remove more than violence, pornography and hate (Gillespie, 2018). Google also installed a feedback system for its web search to report inappropriate autosuggestions, after reports of queries for the 'holocaust' autocompleting with 'is a hoax' (Solon and Levin, 2016; Hern, 2017).

As with new media technologies of old, social media platforms currently are said to enable the 'supercharging' or the acceleration of the spread of fake news (Bounegru et al., 2018). Two terms have been used to capture the web and subsequently social media as accelerationist media: clickbait and computational propaganda. Clickbait connotes titillating and sensational content and is formulaic in its presentation, often containing numbered lists (sometimes referred to as a 'listicle') as well as a cliff-hanger or 'information gap' that sparks curiosity, e.g., 'twenty things you should not do when visiting Japan'. Facebook, in seeking to identify and downgrade clickbait in its news feed, defines it as 'a posted link with a headline that encourages people to click to see more, without telling them much information about what they will see' (O'Donovan, 2018). Generally social media companies seek to operationalize substantive definitions into computational signals. Thus, to Facebook, brief attention (or short 'time-on-site') is a signal of clickbait, for readers, having been lured in to the 'junk food of content consumption', are subsequently dissatisfied with the low-quality content, and leave the page quickly (DeAmicis, 2014). Clickbait, often innocuous, can be combined with divisive content (Burger and Schenk, 2019). 'Extreme clickbait' was a part of the story behind the allegedly apolitical Macedonian teens based in Veles, who used 'spammy techniques' in optimizing pro-Trump sites to make money, in the run-up to the US presidential elections of 2016 (Silverman and Alexander, 2016). Follow-up reporting has sought to debunk that narrative, finding that the clickbait campaign was orchestrated by political operatives (Wendling, 2018; Silverman et al., 2018).

Computational propaganda, the second term, refers to 'the assemblage of social media, autonomous agents and algorithms tasked with the manipulation of opinion' (Neudert, 2017: 3). The breadth of the definition is intended to capture the bots that amplify content, the advertising platforms that enable micro-targeting and personalization of influence messaging, and the click farms that inflate the follower counts and engagement scores, granting posts higher 'vanity metrics' and thus greater symbolic power through fake support (Rogers, 2018a). For computational propaganda, bots increase the spread or reach of the posts and inflate their metric counts

Amsterdam University Press

(Woolley and Howard, 2016). 'Low-credibility content' is spread disproportionately by 'social bots,' which refer to bots or autonomous agents tasked with influencing discussion and public opinion; such a finding has led to calls for curtailing their use (Shoa et al., 2018). As a part of the 'assemblage' of actors and software practicing computational propaganda, the work of software-assisted, political operatives has come under scrutiny, especially in the run-up to elections. Sock puppets, assuming the false identity of a grassroots organizer or a concerned individual, create and circulate political content, organize events and mobilize audiences, making interventions in the physical world through hashtags, internet memes and Facebook events (Mina, 2019). 'Front groups' or even faux 'hashtag publics' also mobilize followings and organize demonstrations (see Table 1.1); one notorious case concerned an anti-Islam protest and counter-protest in Houston, Texas, in 2016, where both groups were mobilized by Russian campaigners operating under the names of the Blacktivists and the Heart of Texas, respectively (Shane, 2018).

A related term for disingenuous content insertion for political ends is astroturfing. It is the artificial seeding of newspapers and other content providers with political (or corporate) advertising disguised as genuine citizen concern. Such content is a different category than sponsored political content, where there are regulations that mandate labelling it as 'paid for by' a particular candidate or campaign (Vaidhyanathan, 2017). Nonetheless there have been calls to have 'masked' political content unmasked and marked as sponsored, however much in the case of a pro-Brexit group, Britain's Future, investigative journalists were long not able to unearth the funding source, despite the transparency of its being labelled.

Particular forms of native social media advertising have prompted the calls for further public scrutiny of political ads, and also perhaps an expansion of the definition of such. 'Dark posts' (aka 'promoted posts') on Facebook refer to micro-targeted advertisements, without a referral page anchoring the content for further investigation (Bump, 2017). Used by political operatives, including foreign influence campaigners, in the US in 2014-2017 and beyond, such campaigning tactics assemble 'keyword publics' algorithmically by querying the Facebook advertising platform for words such as 'second amendment' or other pro-gun terminology and sending advertisements to the news feeds of the tens or hundreds of thousands of those users determined to have such an interest (Angwin et al., 2017). These publics are targeted not so much because they are persuadable voters but rather to have them circulate and amplify messaging.

Amsterdam University Press

Table 1.1 Overview of 2016 fake rallies planned and promoted, as listed in the US indictment of 13 Russian nationals concerning foreign election interference

2016 fake rallies planned and promoted		
Date	*Fake rally*	*Location*
25 June	March for Trump	New York
9 July	Support Hillary. Save American Muslims	Washington, D.C.
23 July	Down with Hillary	New York
20 Aug.	Florida goes Trump	Several Florida cities
2 Oct.	Miners for Trump	Several Pennsylvania cities
12 Nov.	Show your support for President-Elect Donald Trump	New York
12 Nov.	Trump is NOT my president	New York
19 Nov.	Charlotte against Trump	Charlotte, N.C.

Source: Parlapiano and Lee (2018)

Apart from particular social media advertising products such as dark posts, other formats have been identified as energizing publics with divisive messages. 'Image macros', also known as memes, are photos with two lines of text, one opening and one closing line, that are a popular format for political messaging on Facebook and have been among the most shared and otherwise most engaged-with content on the platform (Renner, 2017). Indeed, in the data analysis of the most shared posts of the 'fake' (or astroturfing) activist group pages set up by the Russian Internet Research Agency (Blacktivists, United Muslims of America, Being Patriotic, Heart of Texas, Secured Borders and LGBT United), the image macros and other meme material scored particularly well (Chen, 2015; Albright, 2017; Timberg, 2017).

Russian influence campaigning, Russification and the 'hyperpartisan' style

'Dark globalization' is a term put forward by the historian Timothy Snyder to refer to how knowledge of western societal problems provides opportunities to influence campaigners from abroad, or Russia in particular (2018). In the US Snyder refers to the complex of race, gerrymandering and the electoral college, and the capacity to target voters in specific geographical areas (such as counties in 'swing states') with divisive political messaging that amplify or provide 'oxygen' to viewpoints. There have been detailed analyses of the Russian influence campaign of 2014-2017 commissioned by the US Congress,

Amsterdam University Press

both of which benefited from data provided by Facebook, Twitter and Alphabet (Google) that previously had not been made available for research (Howard et al., 2018; New Knowledge, 2018). They are a part of a litany of literature that has appeared since the commissioning by governments to study the 'tactics' of the influence campaigners as well as the contemporary art of propaganda and the development of counter-narratives more generally. These studies also have led to recommendations concerning how to combat the effects.

The study by the cybersecurity firm, New Knowledge, emphasizes the collective cognitive dissonance that effective propaganda achieves, introducing (and popularizing) language from intelligence and counterintelligence work (2018). Among the goals of the propagandists is to create 'a wilderness of mirrors', originally a phrase from a T.S. Eliot poem but mobilized by the intelligence community (Holzman, 2008). It refers to an environment where truth (and its establishment) are no longer self-evident (Groll, 2018).

To achieve that goal, New Knowledge argues, one particular tactic is the creation of a similarly termed 'media mirage,' or 'interlinked information ecosystems designed to immerse and surround targeted audiences' (2018: 42). They are enveloped in an 'information cacophony', where stories from the press are repurposed, and given another author ('blacktivists'), interpretation and tone. Here is one example, taken from an original newspaper story about how an '11-Year-Old Texas Boy Invents Device to Prevent Hot Car Deaths' (Dahlgren and Arkin, 2017). It was reworked as follows: '[T]hese are stories of Black children the media don't want you to see'; 'White people invent tools for killing, this Black child is inventing a tool for saving lives' (New Knowledge, 2018: 62). The divisiveness and the effectiveness ascribed to the sample post derives not only from the transformation of the feel-good news story into a contrived in-group and out-group divide based on race. Note, too, the format used; the second sentence is a two-liner, to be cast into an image macro or meme, the popular format for sharing and further circulation of grievance, outrage as well as mockery. The story also brings together categories of problematic information. It is both clickbait as well as rather extreme content, and it invites the consumer to read more about the grievance. It is also packaged to be shared.

The purpose of such campaigning is to sow discord and enmity, but it is only one of a variety of tactics where the overall goal is to remove a sense of a collective and shared experience of the world, as analysts have phrased it, and reify group formation (Gessen, 2018). Apart from the creation of a media mirage, the other tactics listed are as follows: 'targeting, asset development, cross-platform brand building, memetics, inflecting a common message for

Amsterdam University Press

different audiences, narrative repetition and dispersal, repurposing and re-titling pages and brands, manipulating journalism, amplify conspiratorial narratives, sow literal division, and dismiss and redirect' (New Knowledge, 2018: 2). With respect to social media, as discussed above, targeting could refer to the audience segmentation available in platforms for advertising purposes, and memetics to the use of both the image macro to formulate a punchy message as well as to build the meme as an additive content container for narrative reinforcement.

It is worthwhile to mention that the expert studies are snapshots, but these as well as subsequent reporting have pointed to the 'ongoing efforts' of the influence campaigners, and their global spread. While social media companies – since the Cambridge Analytica and fake news scandals – have become more active in identifying and suspending accounts of known Russian and other state-sponsored trolls (e.g., Iranian), similarly named accounts are active and can be traced to known networks of political operatives (New Knowledge, 2018; FireEye, 2018). New accounts are continually made (Vaidhyanathan, 2018); the Chief Technology Officer at Facebook speaks of 'blocking more than one million fake accounts every day, sometimes just when they are created' (O'Brien, 2019). The percentage of influence campaigner accounts in that large number is not known.

Recently, there has been growing concern not only about the ongoing efforts of Russian influence campaigners but also the uptake by other groups (or 'domestic actors') of the so-called 'Russian playbook' (Frenkel et al., 2019). Journalistic coverage was prompted by the announcement by Twitter that prior to the US Congressional elections of 2018 it removed accounts of Americans posing as members of state Republican parties (Harvey and Roth, 2018). Facebook also announced that hyperpartisan pages on both sides of the political spectrum in the US would be removed. Discussions of the 'Russification' of online political campaigning also historicized disinformation, pointing to the classic examples, such as the claim that the HIV virus was the leaked product of a US bioweapons lab; it was planted in news outlets beginning in 1983 by Soviet *dezinformatsiya* campaigners in 'Operation Infektion' and ultimately spread four years later to national US TV news (Boghardt, 2009; Ellick and Westbrook, 2018). Comparing the time span of such news spread to the dynamics of reach in the hybrid media system nowadays is how one may describe how the 'platform press' has supercharged fake news (Chadwick, 2013; Bell and Owen, 2017).

In a well-cited article in the *New York Times*, Facebook, as a leading example of the 'platform press', was described as a 'totally insane, unintentionally gigantic, hyperpartisan political-media machine' (Herrman,

Amsterdam University Press

2016). The author spends some time describing the manner in which Facebook mixes posts in its news feed from both family members and faint acquaintances, but also discusses the presence of upstart media organizations and self-styled advocacy groups that only exist online, many only in social media. Most are described as 'hyperpartisan'. These sources populating the platform with content are defined as 'openly ideological web operations' (Herrman, 2016). They also are successful, not just because more extreme and sensational content spreads faster than more sobering truth (Vosoughi et al., 2018). It is also because they employ formats that engage large numbers of users and learn from their engagement and reach. 'Operating' in a continuous feedback loop of metrics data, posts are optimized to perform well in social media. The performance measures are based on the virality of posts, and those that work well are emulated. There are particular formats as well as styles that drive engagement. Memes and clickbait such as listicles, cliff-hanger headlines and human-interest stories are among the formats used, as mentioned above. The hyperpartisan style has a variety of substantive features, not all of which are equally applied, but many appear to be working well. Often anti-establishment as well as positioned as against or in competition with the truth-seeking and fact-finding of the mainstream media, the media operations post stories that are alternatives. These alternatives may be interpretations, facts and editorial commentary on events. They become media layers on the news. The presentation is often edgy, both in terms of being knowledgeably on trend but also sharp in tone. The posts are regular, and as such are part of the permanent updating culture, providing a competing 'feed' about what is happening in the world and in media.

The post-truth condition

There is a series of contemporary utterances that have contributed to public discourse about a post-truth condition. One is the satirical notion of 'truthiness' (Colbert Report, 2005). Developed as political news commentary and comedy, it refers to having the appearance of being true, but without evidentiary basis. Another – 'alternative facts' – is a term that initially referred to the insistence by a member of the US Trump administration that the number of attendees at the presidential inauguration in 2016 was higher than reported and measured by crowd science (Still, 2017). The subsequent clarification of the meaning behind 'alternative facts' is more to the point: 'additional facts, alternative interpretation' (Nuzzi, 2017). Compared to

Amsterdam University Press

truthiness, here facticity does not derive from eye-witnessing or additional methodological authority but rather from other fact-making.

In response to what is sometimes considered first-order objectivity battles, or disputes over matters of fact (Margolis, 1995; Latour, 2008), newspaper reporting with such headlines as 'Here Are the Real [Facts]' as well as the work by fact-checking bureaus and initiatives are contesting fact claims with increasing urgency (Fandos, 2017). These are public debates about facts, inputs into which include fact-checking, a common practice of journalists and university research groups seeking to confirm the basis behind particular statements by politicians and others (Graves, 2016). Recently, scholarship on the effectiveness of fact-checking has developed in at least two directions: the extent to which fact-checking corrects the record as well as factual beliefs, and whether it changes attitudes (Barrera et al., 2017). Both are part of the decades-long discussion and critique of the 'information deficit' and 'diffusion' models, which challenge ideas that providing correctives clears up controversies (Wynne, 1991; McNeil, 2013).

In the fake news scholarly discourse, it has been found that there are distinct audiences for 'alternative facts' and 'fact-checked facts' (Bounegru et al., 2018). Whilst there may be a correction to the record, the original audience may not have been exposed to it. Fact-checked stories also have similar circulation patterns to alternative facts; they are forwarded to like-minded audiences (Shin and Thorson, 2017). Though it does not tell the entire story about exposure, both the original as well as the fact-checking publications are outlets with distinctive audiences or subscriber bases, with fact-checking newsletters often with smaller, specialty circulations, though their visibility may increase as they are built into platform interfaces such as Facebook's. In the other strand of work, it is asked, does exposure to fact-checked facts change factual beliefs as well as attitudes? Here one set of findings is in keeping with the critiques of the effectiveness of fact-checking and the information deficit model more generally, for respondents saw their factual accuracy improve, but their attitudes remain unchanged (Nyhan et al., 2019). Fact-checking, however, could be understood as a documenting process that corrects the record by capturing a dubious story and committing it, and its debunking or exposure, to searchable databases and other media.

The post-truth condition, though, has been described as a competition with respect to not first-order but second-order objectivity. In such a circumstance there is a rise of competing regimes of truth (Fuller, 2018). Expertise becomes 'sectarian' (Turner, 2001). The idea of the media mirage (evoked to describe effective disinformation campaigns) does not in itself create a competing truth regime or infrastructure. Rather, it introduces noise into an infrastructure.

Amsterdam University Press

But when propagandists, or in a different reading of the contemporary situation, a populist radical right media ecology, create an alternative news and information infrastructure, those efforts fit with descriptions of the post-truth condition (Benkler et al., 2017; Sängerlaub et al., 2017).

In other words, post-truth is a term that should not be construed as signifying hoodwinked (or radicalized) consumers, or the 'wholesale cheapening' of fact-making (Sismondo, 2017). Rather, in asking whether 'we can have our facts back', the debate concerns whether (or when) publics can agree on the 'facticity infrastructure' or even the modernist project of knowledge institutions (Marres, 2018). As a case in point, there are ideologically distinctive alternatives to Wikipedia (such as Infogalactic, Metapedia and Conservapedia), producing encyclopaedias challenging not only what is known or settled fact, but also the sources rooting it (Fitts, 2017).

Elections, disinformation, and the Dutch case

Three recurring topics are often discussed in the news and (commissioned) research on disinformation and fake news in the Dutch context. First of all, of particular concern are Russian trolls and their spreading of disinformation in the Netherlands. Secondly, there are the (non-Russian) fake accounts and fake fans that that inflate the popularity of a campaign or a prominent figure, granting them greater symbolic power. And thirdly, publications are addressing its discernibility and possible countermeasures. How to recognize it and combat it? Each of these discussions is often set against the backdrop of a changing news media landscape, whereby mainstream news is increasingly competing with more tendentious and hyperpartisan outlets, and digitization is leading to user-driven and algorithm-driven personalization. That may narrow the horizon of news that users encounter and perhaps increase fringe consumption, though in empirical studies such has not been found (Wieringa et al., 2017). Comparisons of the Dutch situation are also drawn with the US.

While digitization may be changing how people consume news, a study of online news behaviour, disinformation, and personalization of the news by the Rathenau Institute stresses that in the Netherlands, the traditional news media still hold a firm and stable position in the media landscape (van Keulen et al., 2018). The study also finds that there is not (yet) widespread algorithmic personalization in Dutch media sites. And, in stark contrast to the current situation in the US, Dutch news consumers tend to use a variety of sources and have trust in the traditional news media (and less so in social

Amsterdam University Press

media). Lastly, the report underlines that the Netherlands does not have such a particularly polarized media landscape as the US.

Overall, there is a strikingly moderate tone of voice in the literature on the Dutch case, both in news reporting and research reports. Since 2016, several studies have looked at disinformation practices in the Dutch political landscape, and each of them has concluded that neither is there any large-scale disinformation activity in the Dutch media nor does disinformation have a significant impact on Dutch citizens. However, in the Summer of 2017, Wilfred Rietdijk, a Dutch general and national security advisor, announced in an interview with Dutch newspaper *de Volkskrant* that the Netherlands could no longer deal with the digital threat (van Zijl and Modderkolk, 2017). A 'landslide of fake news', as the subsequent tabloid headline read, would lead the country into chaos and division (Jonker, 2017). Including breaches and intrusions in his threat assessment (thereby widening the scope beyond disinformation), Rietdijk explained how Dutch companies are 'in the line of fire' from 'thousands of hackers from Russia, China, and countries such as Iran and even Sudan' (van Zijl and Modderkolk, 2017). The general is not the first to warn of foreign interference in the Dutch online space, though case studies were lacking, at least in the public domain.

Russian trolling and its perceived insignificance in the Netherlands

When the Minister of Internal Affairs, Kajsa Ollongren, warned the Dutch government of Russian disinformation in the Netherlands, she initially was criticized for not having compelling examples (Pleijter, 2017; Kist and Wassens, 2018a). Two journalistic studies that have looked into Russian tweets have found activity in the Dutch online realm, however. A study by *NRC Handelsblad* mined 200,000 tweets from Russian Internet Research Agency (IRA) accounts and found disinformation campaigning beginning in 2014 and another spate in 2016. The weekly magazine *De Groene Amsterdammer* combined the *NRC Handelsblad* data with larger collections of Russian troll accounts, made available on the American public opinion analysis website, *FiveThirtyEight* as well as the lists published by American Congress (van der Noordaa and van de Ven, 2018a). Both studies found a peak in trolling activity after the downing of MH17 in July of 2014. The *NRC Handelsblad* study finds that Russian trolls posted 57,500 tweets, most of which were in Russian and aimed to influence public opinion in Russia and Ukraine, and only four of the tweets were in Dutch (Kist and Wassens, 2018b). The study by *De Groene Amsterdammer* confirms that most tweets on MH17

Amsterdam University Press

were in Russian but finds more mentions of Dutch 'conspiracy theorists and activists', indicating a shift from challenging Western narratives (for Russian-speaking audiences) to seeking to stir conflict within the West.

A second event revealed more coordinated Russian troll activity in the Dutch language Twitter space (in Belgium and the Netherlands), and a further example of striving to foment unrest, albeit unsuccessfully (according to engagement measures) (van der Noordaa and van de Ven, 2018b). It concerned the spreading of anti-Islam content directly following the terrorist attacks in the Brussels airport and metro in March 2016, and in the two years after the attacks. This anti-Islam 'campaign' involved about 950 tweets in the Dutch language that were circulated by some 150 IRA-related accounts. These tweets were rarely retweeted, however. In the event, Russian trolls are more successful in the Netherlands with the circulation of English-language content. While these tweets are not related to Dutch issues and focus on for instance the US elections, they have been shared widely by over 6,000 Dutch Twitter users with a total of 9.5 million followers (Kist and Wassens, 2018a).

Perhaps counterintuitively, there was only minimal Russian interference with the Ukraine referendum in the Netherlands in April of 2016 (NOS, 2017). There was the Russian video capturing fake Ukrainian far-right militia members threatening terrorist attacks in the Netherlands and burning a Dutch flag, but it was readily recognized as propaganda (Bellingcat, 2016). Otherwise, only a handful of tweets propagating a 'No' vote was found in the larger set of tweets under study (van der Noordaa and van de Ven, 2018a).

The *NRC Handelsblad* concludes its work on the Twitter data set by noting that it is possible there is larger scale Russian activity in the Netherlands; it should be studied beyond just Twitter to include other platforms with known troll activity, such as Facebook, Instagram, YouTube and Reddit. Indeed, especially after Trump's victory in the US presidential elections of 2016, many news outlets pointed towards Facebook. As discussed in some detail below, a study by *BuzzFeed News* compiled the most engaged-with posts in the nine months prior to the elections and found that so-called fake news during that time was circulating more than mainstream news. Journalists from the *NRC Handelsblad* replicated the study's general method for the Netherlands, but with a narrower definition of fake news. They determined that the one hundred most-shared political news articles from January and February of 2017, in the run-up to the Dutch general elections, did not contain fake news (Kist and Zantingh, 2017). Certain articles could be considered misleading or biased, they thought, for they exaggerated news facts or took them out of context. The themes that were most resonant during the campaign period in the Netherlands were immigration, Islam and Geert Wilders.

Amsterdam University Press

Dutch fake followers and trolls

Until November of 2017 much of the reporting has insisted that the Netherlands – and the Dutch elections in particular – have been largely unaffected by disinformation or fake news. Much of the news coverage that speaks of it concerns 'fake followers'. For instance, in 2015, there was a small scandal about Geert Wilders concerning a dubious increase in his followers on Twitter. Indeed, when Twitter addressed the issue of fake followers and follower count inflation through a mass removal of suspect accounts in 2018, Wilders as well as other Dutch politicians (including from the political party Denk) saw their metrics decline (NOS, 2018). In perhaps the most well-known case, the Dutch singer-songwriter Dotan was found to have a fake following of 140 user accounts, which were used between 2011 and 2017 to like the musician on social media, edit the Wikipedia article on the artist, request his songs at radio stations and circulate heart-warming stories about him across social media platforms. One of the profiles declared how Dotan's music helped her through a period of grief after a miscarriage; another tells how Dotan welcomed one fan's terminally ill brother in a meet-and-greet, throughout which the singer held the boy's hand. Both testimonials were false, as reporters of *de Volkskrant* found and Dotan later confirmed (Misérus and van der Noordaa, 2018a; 2018b).

In 2018 the first large-scale global study of computational propaganda was published, examining organized social media manipulation such as the use of fake followers in 48 countries, including the Netherlands (Bradshaw and Howard, 2018). The study describes the different computational tactics employed not so much by Russian influence campaigners but by political parties to influence voters and the elections.[2] It was found that the use of social media as an infrastructure for the spread of propaganda and disinformation has become widespread. Under examination is 'cyber troop activity,' defined as 'government or political party use of social media to manipulate public opinion' (Bradshaw and Howard, 2018: 9).

While in more authoritarian regimes, social media manipulation fits into larger scheme of voter suppression and election rigging, in 'emerging

2 The research conducted a content analysis of news articles reporting on cyber troop activity in a sample of 48 countries, supplemented by an in-depth secondary literature review. To collect the news articles, the researchers used the following keywords in combination, in queries across Google, Yahoo!, Bing and LexisNexis: astroturf*; bot; Cambridge Analytica; Facebook; fake; fake account; disinformation; government; information warfare; intelligent agent; military; misinformation; persona management; pro-government; propaganda; psychological operations; psyops; social media; sock puppet*; troll*; Twitter (2018: 8).

Amsterdam University Press

and Western democracies, sophisticated data analytics, and political bots are being used to poison the information environment, promote scepticism and distrust, polarize voting constituencies, and undermine the integrity of democratic processes' (Bradshaw and Howard, 2018: 5). The tactics described include the use of three kinds of fake accounts. First, there is the creation of online commentator accounts that attack and troll genuine users, spread divisive content, or '[divert] conversations or criticism away from important issues' (Bradshaw and Howard, 2018: 11). A second tactic entails automated accounts or political bots to automatedly flood particular hashtags, and astroturf by faking a follower base. The bots also troll genuine users by reporting them and flag organic content thereby having both suspended until a human moderator checks them. A third tactic is the use of hybrid accounts, which are those that make use of automation (for the sake of speed and convenience) but are actively curated by human users, who commonly manage multiple fake accounts or sock puppets. This type of fake account is difficult to recognize, and thus to combat. The study finds that automation is the most recurring tactic, seen in 38 of the 48 countries under study.

Besides fake accounts, other strategies involve the use of political ads and the involvement of search engine optimization and activity on chat applications and across social media platforms. Where Twitter is proven to be the platform most friendly to automation, the study finds 'cyber troop activity on chat applications or other platforms (Instagram, LINE, SnapChat, Telegram, Tinder, WeChat, WhatsApp)' in one-quarter of the countries under study (Bradshaw and Howard, 2018: 13). In the European countries in their sample, they find distinct junk news footprints per country. In Germany, it is rather marginal and was mostly circulated by far-right political actors during the 2017 federal elections. In Italy on the other hand, a large and active 'ecosystem' of it is connected to political forces such as the Lega Nord (Northern League) and the Movimento Cinque Stelle (M5S, 5 Stars Movement), which were at work during the 2017 constitutional referendum and the elections of 2018. Here, junk news connects national politics to Euroscepticism, conspiracy theory, aliens and pro-Putin propaganda. In the Netherlands, the analysis finds that it revolves around politician Geert Wilders and in particular the spread of his anti-Islam video, which was broadcast on television and shared in social media in the lead-up to the 2017 Dutch national elections. In particular, the study finds that automated accounts have amplified Geert Wilders' campaign hashtags.

These results match the findings in a study that looked at troll-like behaviour on Twitter, leading up to the 2017 Dutch general elections, where sock puppets were found (Bounegru et al., 2018). The study collected over

Amsterdam University Press

500,000 tweets mentioning at least one of the Twitter accounts of the 28 political leaders a month before the 2017 Dutch general elections. To retain the users that demonstrated troll-like behaviour, it narrowed down the set to only the 25 users who consistently targeted one or more political representative.[3] The analysis showed that there was a notable asymmetry in the distribution of targets of troll-like behaviour and sock puppetry across the political spectrum, where left-wing politicians are most often targeted by negative mentions, while right-wing politicians receive support. Troll content extended to reputable news sources which cited it at least thirty times. Among the cited troll accounts were fake news organizations with names as 'Today in Syria' and 'WorldNewsPolitics', political parties (including multiple fake accounts for the Republican party in Tennessee) and concerned citizens, most of whom were fiercely pro-Trump and anti-Islam (Kist and Wassens, 2017). In another analysis by the *NRC Handelsblad*, a Dutch political party (DENK) also exhibited troll-like behaviour, including sock puppetry on both Twitter as well as Facebook (Kouwenhoven and Logtenberg, 2017).

While Dutch news consumers have been found to use a variety of news sources, the Netherlands also has a steady 'pulp news' diet (Burger et al., 2019; van der Poel, 2019). From 2013-2017 Dutch Facebook users consumed more low-quality, commercially driven clickbait than mainstream news, as was found through engagement scores. As may be expected, there is also relatively more clickbait on Facebook than quality news.

The consumption and forwarding of clickbait, extreme clickbait as well as other problematic information extends also to politicians and public figures. One Dutch researcher, Peter Burger, has a collection of instances when Dutch politicians have retweeted anti-Semitic or otherwise disturbing content. In one example, a video purporting to show 'Muslims vandalizing Christmas market in Lithuania' was actually a recording of an event that took place in the city of Baltimore in the US (Burger, 2016).

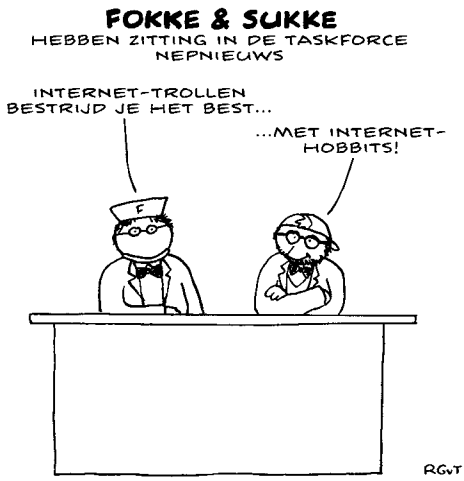
Recognizing and countering disinformation in the Dutch online space

Various initiatives aim to detect and counter disinformation in the Netherlands and on an EU-level. The EU taskforce (*East Stratcom Task Force*) against disinformation was heavily criticized in the Netherlands after its project EUvsDisInfo mistakenly categorized articles by *The Post Online*, *GeenStijl* and *De Gelderlander* as disinformation (van Keulen et al., 2018; Heck, 2018). (Figure 1.1 shows a cartoon about the fake news taskforce,

3 By @mentioning them at least 100 times in a one-month period.

Amsterdam University Press

Figure 1.1 Cartoon that ridicules the fake news taskforce, stating: 'internet trolls are best countered by internet hobbits'



Source: Reid et al. (2018)

stating internet trolls are best countered with internet hobbits.) In a sense, the dispute stemmed from misreadings of single stories, perhaps without an appreciation of how settled some of the sources are in the Dutch media landscape, despite their tendentious style (in the case of *The Post Online* and *GeenStijl*). For its part, *De Gelderlander* had taken over nearly verbatim a Russian storyline concerning the perpetrator behind the downing of the MH17 but did attribute it to its original source in a barebones reporting style. The flagged cases were removed from the EUvsDisInfo site after complaints by the Dutch media organization Persgroep (EUvsDisinfo, 2018).

Fact-checking as a journalistic practice has taken hold in the Netherlands. Many newspapers have implemented (or revived) a fact-checking section, often dedicated to checking statements made by political figures in interviews in newspapers or TV shows. There are also websites such as Hoaxmelding.nl and Nieuwscheckers.nl that compile lists of instances of false news on Facebook and elsewhere. For their study of disinformation, Rathenau researchers analyzed these lists, comprising respectively 140 on Hoaxmelding (collected between 1 February 2014 and 18 December 2017) and 166 on Nieuwscheckers (between 3 February 2017 and 5 January 2018) (van Keulen et al., 2018). They found that the items on the list of Hoaxmelding involved examples of unfounded warnings (65), polarizing disinformation (32) and fake crime news (31). Additionally, there were several examples of clickbait, benign as well as malicious. The content steers users to

Amsterdam University Press

Figure 1.2 'Detected and eliminated' fake news, with a warning issued by NU.nl and Nieuwscheckers

Source: NOS (2017a)

advertising, 'like-farming' and phishing sites (van Keulen et al., 2018: 38). Such posts contain human interest stories that are 'painful on a personal level' (van Keulen et al., 2018: 45). The researchers found that only 25% of the disinformation concerned political content and most clickbait serves a commercial goal, rather than a political one. On the list of items collected by Nieuwscheckers, the Leiden University-based initiative, less than half was found to have political content. Within the full set, the researchers found six examples of polarizing content. Otherwise, many of the posts concern factually incorrect, public statements by politicians, the investigation of which is how fact-checking is conventionally practiced.

Fact-checking now extends well beyond unpacking politicians' statements, and Facebook has entered into partnerships with many bureaus around the world, including in the Netherlands, to explore and catalogue dubious content. In 2017 Nieuwscheckers partnered with Facebook and NU.nl and celebrated their first collaborative, 'successful detection and elimination of fake news' that year when they flagged a tabloid-style, human-interest post about an Australian new-born weighing 20 kilograms (see Figure 1.2). In February of 2019, however, Nieuwscheckers withdrew from the Facebook fact-checking initiative because of liability risks (Kist, 2019). Nu.nl continued to work with Facebook on fact-checking, on a paid basis, an issue raised

Amsterdam University Press

repeatedly in the context of journalists' being asked to address an issue of Facebook's making on a voluntary basis.

The effectiveness of fact-checking as a strategy in the Netherlands is a different question. As mentioned above, fact-checks and fake news often have separate publics, and fact-checks may lead people to fake news, rather than away from it. A recent study in the Netherlands found that even when many people would agree with a fact-check, they are not interested in reading the fact-checking article, prompting the scholars to advise journalists to make the fact checks an engaging read (Hameleers and van der Meer, 2019). Another strategy to counter disinformation concerns a strand of media literacy that involves developing skills to recognize fake user accounts and disinformation. One is on a source level, the other on a story level. The *Field Guide to Fake News* provides a method for the detection of trolling accounts by looking at their friends, or their profile information (Bounegru et al., 2018). There are also courses and training modules for fake news detection and fact-checking, such as those given by Radio Netherlands (RNTC, 2019). The other format is the fake news quiz, such as those by *de Volkskrant* (2016) and the *Guardian* (2016), as well as the *New York Times* 'deceptive Facebook post' test (2018). These quizzes make it clear how challenging it is to recognize fake news. The Dutch serious game, titled *Slecht Nieuws* ('Bad News'), invites players to create fake news and by doing so gain insight into the strategies behind it and become more astute in its recognition (NRC, 2018; DROG, 2018). It is part of efforts that study false news as risk and ultimately seek to inoculate populations against it (Roozenbeek and van der Linden, 2018).

Voting aid applications

Voting aid applications (VAAs), often called *stemwijzers* in Dutch, are generally websites that describe their purpose as helping undecided voters find the political party that best matches their preferences and positions. As such, in the context of the study of disinformation and so-called fake news, they could be regarded as a competing persuasion instrument, a pre-emptive measure against influence campaigning, or even a potential site that may include it, either through parody, hoax or hack. The literature on VAAs takes up the Dutch and Belgian cases, countries that together with Germany, Austria and Switzerland have upwards of half the voter population accessing them prior to elections. The work can be positioned broadly as pertaining to 'the impact of internet-based applications on politics' and can be roughly divided into user studies, impacts of VAAs on the voters as well as the methods behind them (Hirzalla and van Zoonen, 2015: 88). To date

Amsterdam University Press