

Of Trolls & Bots

*The Basics
of Manipulation in
Social Networks*



Co-funded by the
Erasmus+ Programme
of the European Union

Imprint

Project Leadership

Dr. Sebastian Fischer
Institut für Didaktik der Demokratie
Leibniz Universität Hannover

Project Management

Arne Schrader

Authors

DETECT-Consortium

Design

Mareike Heldt



Co-funded by the
Erasmus+ Programme
of the European Union

Copyright:



All rights reserved. The content of the publication may be used for educational and other non-commercial purposes on the condition of using the following name as source in every reproduction: «Erasmus+-Projekt DETECT».

Project Website:

www.detect-erasmus.eu

This project has been funded with support from the European Commission. This communication reflects the views only of the author, and the Commission cannot be held responsible for any use which may be made of the information contained therein. Project Number: 2018-1-DE03-KA201-047400

Consortium

Leibniz Universität Hannover
Institut für Didaktik der Demokratie
Germany



GONG
Croatia



Gimnazija Pula
Croatia



Centre for European Refugees,
Migration and Ethnic Studies,
New Bulgarian University
Bulgaria



Center for Education and Qualification
Bulgaria



Demokratiezentrum Wien
Austria





Contents

Introduction	05
1. Facts, Facts, Fake	09
1.1. What is Fake News?	09
1.2. What is Computational Propaganda?	11
1.3. What is a Conspiracy Theory?	12
1.4. What is a Parody?	13
2. Manipulative Technologies	15
2.1. Social Bots	15
2.2. Trolls	17
2.3. Hoax Campaigns	18
2.4. Algorithms & Filter Bubbles	18
3. Why is False Information so Resilient?	21
3.1. Politically Motivated Reasoning	21
3.2. The Bandwagon Effect	22
3.3. The Mere-Exposure Effect	23
3.4. The Continued-Influence Effect	24
3.5. The Problem with Pictures	25
3.6. When Emotions Meet Algorithms	26
3.7. It's all About the Money	28
4. How to Detect Fake News	29
4.1. Investigating the Content Creator	29
4.2. Investigating the Machine: How to Identify a Social Bot	30
4.3. Investigating a Website	31
4.4. Investigating a Text	31
4.5. Investigating a Picture	33
4.6. Investigating a Video	33
5. Wikipedia – Treat with Caution!	35
6. My Digital Self or How to Be a Conscious User	37
Appendix	40
References	47



Introduction

DETECT aims to improve critical competence of judgement among teachers and students and to strengthen active digital citizenship. As participation is understood as the "fundamental, performative element of citizenship", the digital society faces new challenges and shifts in dealing with citizenship itself by simplifying practices through new media, but also by endangering them (Kenner & Lange, 2018). One dangerous element constitutes the novel negotiation of facts in the digital world: "Facts are weakened in three different, equally powerful ways – political, symbolic, digital. [...]. Facts are weakened by both the rise of populism and the conditions that make possible the populist turn. [...]. One way of countering populism is through citizenship – contestatory, solidary, digital, and creative." (Krasteva, 2017)

Young internet users get informed about the world through platforms such as YouTube, 8chan, reddit and Instagram, in addition to what they hear from parents and peers. Hence, the focused strengthening of active citizenship, through activities in classrooms and on social media, is strongly recommended as an antidote against the abuse of digital media in terms of false information.¹

Therefore, DETECT offers a range of different materials for teachers and students to become aware of false information and to implement learning-processes in their classrooms. As a start, the partner-organizations conducted interviews with teachers in Croatia, Bulgaria, Austria and Germany to analyze their needs in terms of teaching critical digital literacy in schools. Following the results of these conversations, a compendium was developed providing information about general aspects of manipulated content and specific ways of recognizing and resisting it.

¹ When it comes to false or manipulated content, a lot of terms are being used (fake news, computational propaganda, false information). We discuss important terms in chapter 1. In this compendium, we use the term false information. Here, we distinguish between misinformation (false information/content) and disinformation (manipulated and falsified information/content).



*Needs analysis :
European Teachers Reflect on Digital Citizenship Education*

Today's societies have become very well aware of the reality of manipulated content and the impending danger its rapid spread poses to the fibers of a society's social and political life. Faced with these threats within the context of a democratic society, it is pertinent to ask how the competencies of students can be honed to detect disinformation and in turn promote active digital citizenship. The role of teachers in this process cannot be overemphasised. Teachers of digital literacy should be equipped with the appropriate digital, democratic and didactic know-how. They bear the responsibility to guide students within the framework of inquiry-based learning, and thus enable them to develop practical strategies to identify and understand instruments used for influencing public opinion.

Experiences from Austria show that teachers highlight the importance of promoting media and democracy. This is reflected in the significant number of their expressed interests in organized research workshops. Practical teacher activities also show that the synergy amongst teachers is more productive and positive when they learn in groups, in contrast to working independently. In addition, teachers are aware of the students' dependence on YouTube as a primary source of information to form their opinion. However, the not very critical attitude of students (or total lack of it) in their daily consumption of news and information on social media platforms has not gone unnoticed. Consequently, this uncritical use of information from various internet sources by students has a negative impact on their ability to distinguish between facts and disinformation. Overall, teachers consider the absence of classroom discussion on this topic, the missing systematic methodology on how to teach students to identify manipulated content, and the time-consuming intensive planning as challenging.

Respondents in Germany report a slightly different learning experience. For one teacher, an inquiry-based learning approach is mostly employed in the classroom during simulation games as a way of engaging with distinct political categories. Another approach is problem-based learning. Through the identification of a problem, a critical learning process is activated, and students are trained to critically engage with an issue, which in turn sparks new questions. According to some respondents who apply this problem-based learning approach, students can "learn about certain entities by answering their [own] question", because "these questions structure learning processes". Although there is the expressed concern that students lack the necessary experience when using these techniques, there is,



however, a consensus amongst teachers about students' interest in inquiry-based and problem-based learning. Similar to the Austrian assessment, respondents in Germany find these projects to be generally time-consuming. This, in turn, has its effect on other subjects. Consequently, teachers might not immediately recognize and embrace the importance of this topic. Furthermore, all teachers acknowledge that they do not have systematic experiences with this topic in their classes. It is not included in an already too comprehensive curriculum and there is no time to incorporate manipulated content as an additional topic to be covered.

Similar sentiments have been expressed by their Croatian colleagues. Based on their experience, they stress the genuine threat faced by students who are very often not able to recognize the dangers of the internet when exposed to the abundance of online information. This deficiency makes students not resilient enough to forms of disinformation. A well-developed media literacy competence, as well as a critical outlook, is of utmost necessity – both for the students and teachers. For instance, many students use Wikipedia as their first and only source during online research. While public polls confirm these information literacy deficiencies in a considerable number of youths, they also show us the substantial help and guidance these students need. To address this gap, teachers rated extracurricular activities as the most significant contributor to developing media literacy, closely followed by singular projects and “project days”, a form of participation practiced in schools. They also admitted having employed the use of different formats in the course of implementing media literacy activities at different points in time. Lastly, the most conspicuous challenge faced by Croatian teachers is the missing framework to support them, as well as the inflexibility of formal education formats within media literacy that can be examined and studied.

The experiences in Bulgaria draw a somewhat different picture. Respondents report of dialogues and discussions as teaching tools in class and during projects, seminars, and extracurricular activities. For these teachers, DETECT-studios provide an innovative and useful platform to improve digital literacy, civic education and to prepare students to become digital citizens. Teachers report the level of digital literacy in Bulgaria as low. This limits the students' ability to differentiate between facts and disinformation. However, there is a significant readiness to develop the needed skills to overcome media literacy deficiency. For this reason, Bulgarian teachers are enthusiastic about participating in DETECT-studios as it helps them to embrace best practices and creative tools which in turn could be deployed in schools, especially when teaching different subjects such as philosophy, history, and psychology.



By way of illustration, with regards to inquiry-based learning and the idea of a DETECT-studio, a teacher in entrepreneurship asserts that “she practices all the time inquiry-based learning and she does not teach otherwise”, because it helps “students achieve very good results [and] strengthen a variety of competencies [...]. Students are also taught how to learn and realize the importance of life-long learning”. These acquired competencies are critical for young students as they help them to navigate and orientate themselves in the information flow. One drawback, however, remains, as “several teachers are more conservative and prefer traditional methods, some teachers are afraid of experimenting”. In the same vein, some other teachers, for instance, in the field of literature and language, business ethic and business communication, information technology and history, as well as philosophy have attested to putting to use inquiry-based learning. For them, possessing digital competencies, which constitute an element of media literacy, is of great value in contemporary society. To that effect, DETECT-studios offer an enriching framework for enhancing these proficiencies.²

As shown, teachers need basic information on forms of disinformation and manipulative technologies on social media. This handbook gives an overview of important terms, strategies and instruments, providing basic information to both recognize and resist manipulated content.

² For more information about the conducted interviews please visit www.detect-erasmus.eu

Facts, Facts, Fake

Chapter One

Fake news, computational propaganda, conspiracy theories, misinformation vs. disinformation, etc. When it comes to false or manipulated content a lot of terms are being used. As part of (digital) media literacy, it is essential to know what you are talking about and why there is a difference between poor journalism and political propaganda. In this chapter, we highlight and discuss important terms that will help you navigate the universe of false or manipulated content.

1.1. *What is Fake News?*

The European commission defines 'fake news', 'post-facts' and 'alternative facts' as terms that "refer to perceived and deliberate distortions of news with the intention to affect the political landscape and to exacerbate divisions in society" (European Commission, 2018). In the JRC Digital Economy Working Paper 2018-02 two definitions of fake news are presented:



A narrow definition would be limited to verifiably false information. Fact-checking can expose false news items and identify the sources of these articles. Most empirical social science research on fake news follows this narrow definition because it requires an identifiable and well-defined set of false news articles and sources to measure the reach and impact of false news (Alcott & Gentzkow, 2017; Fletcher et al, 2018). Some measures taken by social media networks against fake news concentrate on verifiably false news: hiring fact-checkers, tagging suspicious postings, removing false news posts, etc. [...].

A broader definition of fake news would encompass deliberate attempts at disinformation and distortion of news (European Commission, 2018a; Wardle & Derakshan, 2017; Gelfert, 2018), the use of filtered versions to promote ideologies, confuse, sow discontent and create polarization. [...].





A more differentiated definition of fake news was coined by Claire Wardle, Harvard professor and founder of First Draft, an organization that is devoted to tackle the issue of manipulated content. In her work, Wardle initially distinguishes between misinformation and disinformation. Whereas misinformation describes false information that is unintentionally created, disinformation is made and disseminated intentionally. According to her, “disinformation is false information that is deliberately created or disseminated with the express purpose to cause harm. Producers of disinformation typically have political, financial, psychological or social motivations” (Wardle 2018).

Although both forms of information are problematic, this distinction is necessary. False information will always occur, even when the highest quality of journalistic standards is applied. False information/misinformation must be retracted, and systems must be created to avoid or reduce them. However, it is especially strategic organised disinformation campaigns by right-wing groups or conspiracy theorists that are a danger to democracies. They are designed to stir-up hate against marginalised groups and to spark distrust in democratic institutions and processes (i.e. elections), scientific knowledge and critical media (Wardle, 2017).

According to Wardle (2017), the next step then is to ask why a specific content was created in the first place in order to distinguish between misinformation and disinformation. For this, she defines fake news as a spectrum organized from – among others – poor journalism to propaganda. To locate false information on this spectrum it can be helpful to question the content’s intention. For instance, was this article written to inform me about contemporary events, yet failed to get all the facts straight (poor journalism)? Was this video manipulated to make me laugh (parody)? Was this comment written to provoke (provocation)? Was this video produced to create profit (profit)? Was this graph manipulated to increase political influence (power)? Was this entire article fabricated to make me question democracy and stir-up hate against marginalised groups (propaganda)?

This also affects acts of “giving voice” to topics without major consideration, which are now included and reported in traditional media once they gain prominence online. Obviously, this can be a good thing if the “voice” is e.g. given to marginalized groups, but more often, it is a problem of online sensationalism that becomes so widely popular that professional media seamlessly takes it on.

It is also important to accept the term post-truth in connection with fake news and fake content.

John Corner (2017) said that “post-truth” and “fake news” are qualified as key indicators in analysing the current media and political situation, and the focus of research is on numerous reviews in newspapers and magazines as well as in new media. The author points out that there is unpredictability and uncertainty in the public dissemination of facts and “truth”. This emphasizes that when creating news, principles must be observed and respected, and there should be a measure to prevent the deliberate falsification of information that is not a good journalistic practice or is the result of fraud strategies and unprofessional use of sources. Similar plans include other publications looking for cross-points between a poster and fake news. The role of the media and the emergence of various factors in the establishment of the context is presented in the article by Hariklisna Bashharan, Harsh Mishra and Prader Nair (2017), who have expanded on the distribution of the fortune and talked about the era of popularity, where there is room for fake news.

To understand computational propaganda, it is also necessary to understand the context of disinformation, the narratives in which they tap into, the political and social cultures in which they are produced and spread (Bounegru et al, 2018, 8). For instance, right-wing groups create disinformation for three reasons: 1, to target marginalised groups, 2, to spark mistrust against the state and the media, while simultaneously 3, providing “alternative” news and facts. Furthermore, the larger its reach, the more impact it will have. Not every false content will gain a wide range. However, strategically planned networks of websites, blogs, Facebook pages, etc. help to accelerate this process.

1.2. What is Computational Propaganda?

In general, 'propaganda' refers to true or false information that is disseminated to persuade an audience for political purposes. It is part of a larger group of deliberate information campaigns that can be described as “advertising, public relations, public diplomacy (or public affairs), information operations” (Jack, n.d., 4). Since persuasive information campaigns usually blend facts and interpretations, it is difficult to assess their accuracy. In effect, the labelling of a campaign as publicity or propaganda can substantially rely on the perspective of the observer (4). In many languages, e.g. Spanish, 'propaganda' is used for both the concepts of publicity and persuasion. However, in English, 'propaganda' is, in most cases, a pejorative term that implies the intent to manipulate or deceive. The design of propaganda can cultivate attitudes of the audience and/or provoke action (7). During the early twentieth century, propaganda had neutral or positive connotations for some scholars. However, the term became



negatively associated since the German Nazi Party's adoption of the word for one of their ministries and its anti-Semitic propagandist campaigns (8).

Until today a critical reflection of propaganda remains necessary. Political events in recent years led to an increased interest in the concept of propaganda. News stories are created by (political) entities to influence public opinion. Sometimes based on facts, these news are always biased, since they favour a specific point of view. While appearing as objective pieces of information, the purpose is not to inform but to persuade the public (Tandoc et al., 2017, 146-147). Additionally, it is important to note that propaganda strategies are not only used by governments and political parties but can be applied to the actions of both governmental and non-governmental actors to criticise them (Jack 8).

Although propaganda itself is not a new concept, with the introduction of the social web its means of dissemination has changed. Present political propaganda operates with new forms of technology such as social bots. Samuel Woolley (2016), director of research of the Computational Propaganda project in Oxford, claims that political actors around the world have increasingly made use of the digital power of social bots to manipulate public opinion and disturb communication. These software programs mimic human users on social media platforms. Politicised social bots are used in several ways: they boost politicians' follower levels to generate impressions of popularity and they flood news streams with spam or send out sophisticatedly manipulated information. This distribution of disinformation with new forms of automatization is what we call 'computational propaganda' as defined by Samuel Woolley & Philip Howard (2019). The term computational propaganda is useful as it describes this specific combination of technical aspects (algorithms, social bots, etc.) as well as societal aspects.

1.3. What is a Conspiracy Theory?

The internet frequently provides its users with 'alternative explanations' that hold 'evil forces' responsible for a broad range of issues such as 9/11, the missing cure for cancer or the 'refugee crisis'. It is safe to say that there exists a conspiracy theory on the internet for almost every complicated process or event that is not that easy to understand (Hummel, 2018, 187). The history of anti-Semitism illustrates that conspiracy theories "can have very real consequences and are capable of creating a highly disquieting social reality" (Heins 2007, 788).

A conspiracy theory often consists of three basic components. First, a collective of conspirators; second, the existence of a plan the collective follows; third, the secret implementation of the plan. Moreover, the dualism of good and evil is of importance: the conspirators' actions apparently harm other people. Another characteristic of conspiracy theories is that they contradict the "official" version of events. The emergence of conspiracy theories can be explained psychologically: By constructing links between events – and in this way explaining them – conspiracy theories provide their believers with security. They offer 'special' knowledge and supposedly protect their supporters of harmful influences (Hummel, 2018, 188-189). Furthermore, the belief in conspiracy theories has often been traced back to societal crises situations. In these situations, feelings of fear, uncertainty, and the belief of being out of control frequently emerge. Situations of uncertainty stimulate the desire to comprehend the environment (Prooijen & Douglas, 2017, 329). The "building" of identities around narratives is also important, as it is exploited in such a way that we are presented with "enemies" endangering our identity, ergo our existence. Opposite to this dehumanizing process, narration or storytelling can also be put to good use by exposing people to stories that connect us.³

That being said, the mechanisms of the internet still contribute to the cultivation of conspiracy theories in certain ways. Nowadays, everyone can disseminate his or her theories through social networks, blogs, or YouTube videos. The internet also enables believers of conspiracy theories to easily connect with other adherents of conspiracy theories, which may even reinforce their beliefs (Hummel, 2018, 191-192).

1.3. What is a Parody?

The genre of parody has many layers and is consumed often on social media. Parodies should not be considered credible sources. However, it is important to distinguish between their objectives, as some of them can be rather harmful. In general, you can distinguish between two kinds of parody. First, parodies of movies, music videos, famous people and so on, which is a form of ironic or satiric imitation. Second, news parodies (news satire) that inform about and take stance on a political issue but present themselves openly as comedies (i.e. "Gospodari Na Efira", "The Daily Mash", "The Onion"). Some of them, like "Last Week Tonight with John Oliver" or "Neo Magazin Royale" are based on extensive research on social and political issues. Although they, too, inform about and can inspire to think about political issues, their main goal is to make people laugh.

³ Check our webinars for more on this



Besides these harmless forms of parodies, other trends are more troubling. For instance, when the label of parody is misused to spread disinformation. A popular example is a manipulated article created by the site "nachrichten.de.com". This article claims that asylum seekers in Germany received 700 € for Christmas. The article was shared 100.000 times on social media. When you visit the site and scroll to the end of the article, you will see that the whole text was made up and labelled as "parody" (Wolf, 2018). However, as social media expert Ingrid Brodnig (2017) argues, parody must always be recognised as such. In this case, the label was misused for a racist purpose and can be, thus, identified as computational propaganda.

Lastly, the main question here is: when stripped away from jokes and laughter, what message stays behind? The question can be tricky sometimes as the line between a poor joke and hate speech can be thin. This fact is misused when anything is justified by saying: "It's just a joke!".

Manipulative Technologies

Chapter Two

Political propaganda and conspiracy theories are both exploiting false content and the production of disinformation to follow certain aims. As aforementioned, they are a threat to democratic societies by restricting political pluralism and fostering distrust in information and the democratic system. To reach their objectives in social media and digital formats new technologies are used. In this chapter, we discuss the most relevant of these technologies.

2.1. Social Bots

Social bots are omnipresent on social media platforms and throughout the internet. These automated software agents gather information, make decisions, interact with, and imitate real users. Social bots differ from more general web bots as they directly communicate with humans on social media platforms, in the comment section of online news sites, in forums, etc. (Woolley, 2016).

Social bots do not have their own opinion but follow a pre-defined agenda. They aim at connecting with other, real users and build virtual 'friendships'. As soon as the connection is established, real users see when the social bot reacts to contents by commenting, sharing or liking. If real users, in turn, share these contents, all their social media contacts get access to them as well. Due to this snowball principle, the scope of the original post increases drastically (Graber/Lindemann, 2018, 57).

There are different kinds of social bots, one of them are 'fame enhancing bots'. They follow users to increase their popularity and fame. They are most commonly found on Twitter. According to an Oxford University study, "Pro-Leave Twitter bots played a 'strategic role' in EU referendum result" (Sulleyman, 2017). These bots are not only employed to raise follower numbers of politicians and celebrities, but they have also become a common tool for marketing purposes and are thus used to increase the popularity and fame of brands and products (Leistert, 2017, 224).



Programming social bots is not very difficult and can be done by using freeware software. Most social bots work in simple ways: they scan Twitter timelines or Facebook posts for certain words or hashtags and comment on them with prefabricated texts or try to uphold a real conversation, which often proves to be difficult. In some cases, social bots can produce their own answers. These consist of texts or entire statements taken from certain websites. Therefore, social bots' fabricated text will differ. Depending on how well they are programmed, their answers will make sense, at least to some extent. Although they are often misused, social bots themselves are not necessarily malicious. Initially, they were programmed to help people orient themselves on social media or to collect and retweet news items on a certain topic. However, events such as the 2016 U.S. presidential election have shown that these programmes can be utilized for manipulative purposes (Schönleben, 2017).

Furthermore, social bots can become a danger to democracy if they are employed for propaganda purposes. If one software program controls hundreds or thousands of Twitter accounts, it has the power to influence public opinion. For instance, likes and retweets of social bots can manipulate the so-called “trending topics” on social media or on Google search (WerdeDigital, 2016). The perceived credibility of media is not primarily dependent upon the truthfulness of its facts but is highly based on its distribution. Information that shares only questionable bonds with reality may appear truthful if enough people believe in it (see chapter 3.2.). Therefore, even if a media content has been identified as a hoax (see chapter 2.3.), it can have transformed, due to its massive distribution, into the majority opinion. By means of dissemination, social bots can artificially reinforce reports and are effective because of the psychological principle of 'social proof' (Graber/Lindemann 2018 58-59).



© Ka Schmitz

A person programs social bots to spread the same message again and again.

2.2. Trolls

Trolling means to deliberately post inflammatory or offensive content to an online community. The intent is to garner emotional reactions, provoke other readers, disrupt conversation, or silence users. A person harassing, or insulting others online is called a troll. Sometimes the term is used to describe accounts controlled by human performing bot-like activities. A troll farm is an organization or a group of individuals aiming to create conflict by systematically spreading hate on social media. For instance, a troll farm, the Russian Internet Research Agency, is known to have spread offensive and inflammatory content (i.e. against Hillary Clinton) in an attempt to interfere in the U.S. presidential election between Donald Trump and Hillary Clinton 2016 (Wardle, 2017).

2.3. Hoax Campaigns

According to the Oxford English Dictionary (OED) hoaxing is defined as “a humorous or mischievous deception, usually taking the form of a fabrication of something fictitious or erroneous, [...]” For instance, in 2010, a television station in Georgia broadcasted a false announcement. However, what was planned as “a mock half-hour report about a Russian invasion of the country” triggered a national panic (Watson, 2010). Secor and Walsh (2004) explain hoaxing as following:

“ Something is made public, people react, taking it seriously, then somehow the rug is pulled away, and people first suspect, then realize that they have been fooled. Sometimes a state of uncertainty prevails, and the event just fades from public consciousness; sometimes the hoaxer gets unwillingly unmasked much later; sometimes the hoaxer is exposed to public opprobrium [disgrace]; more often, the hoaxer claims credit to construct public notoriety for himself or herself. (ibid.) ”

Some hoaxes are also deliberate propagandist schemes and malicious falsehoods created with the goal of fearmongering, hurting political opponents, and instigating conspiracies. A good example would be the so-called anti-vaccination-movement.⁴

2.4. Algorithms & Filter Bubbles

A computer performs an algorithm, a fixed series of steps, to finish a task or to solve a problem, to categorize and classify. Social media platforms make use of algorithms to compile the content users see. Based on a user's previous engagement on the platform, the algorithms show filtered material according to the user's interest (Wardle, 2018). For instance, the search engine Google provides not neutral but highly personalised search results (Stegemann, 2013). Algorithms are used for various other purposes: They are utilised for personalised advertising, decide whether people can take up a loan, propose which applicants should be invited for a job interview, and are able to predict certain illnesses early on (Schaar 2017).

⁴ More about the Anti-vaccination movement see chapter 3.4. For a broader and more elaborate discussion: Azhar et al (2018). The Anti-vaccination movement: a regression in modern medicine. *Cureus* 10(7). Retrieved from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6122668/>



Other widely known algorithms are used by Facebook, Instagram, YouTube, etc. For instance, the Facebook algorithm filters all the news that could be shown to every user and only displays the most relevant content. Initially, this process is needed because an average Facebook user would otherwise see 400-500 different types of contents every day, and a user with many 'friends' on the platform would even be exposed to a couple of thousand contents every day. Three basic factors are, among other factors, relevant for the newsfeed algorithm. First, affinity measures the quality of relationship between a user and the page owner or content provider to determine how interested a user is in certain contents. Second, weight takes the interactions (likes, shares, comments of the user or his or her friends) into account. Third, decay is concerned with the time decay between the time of the publishing of a post and the last login of the user. If content gets much attention, it will be in the newsfeed although it may be a bit 'older' (AllFacebook, 2016).

On video-platforms like YouTube, the algorithm calculates which video is recommended and proposed to you next, based on your interests and the interests of people from your social network, based on what is popular in your region, etc. Only few people are aware of the hidden dangers. In fact, the algorithm also promotes most of the circulation of conspiracy theories (Lewis, 2018).

The use of algorithms has consequences for the public. Some algorithms subtly modify and amplify media perception (Roese, 2018, 326). This phenomenon can be described with the notion of 'filter bubbles', coined by Eli Pariser. The filter bubble is "a unique universe of information for each of us [...] which fundamentally alters the way we encounter ideas and information" (Pariser, 2011, 9). Although the consumption of media is to a certain degree always based on personal preferences, the filter bubble introduces three new dynamics. First, tailored to individual interest, every internet user has his or her own filter bubble, which automatically separates people. Second, the filter bubble is invisible because the process and the criteria through which sites filter information (how the algorithm was coded) is opaque to the users. From within the bubble, it is almost impossible to notice any bias. Third, while the consumption of traditional media results from an active choice, users are not able to make a choice with personalised filters. These filters approach users and are difficult to avoid (Pariser, 2001, 9-10). In addition, Vivian Roese (2018) claims that through the filter bubble confirmation bias is fostered and the segregation between different groups of people increases because users aggregate in groups of interest (326).



© Ka Schmitz

Not everybody sees the same content due to filter bubbles.

Why is False Information so Resilient?

Chapter Three

False information is resilient for many reasons. First, due to the aforementioned algorithms and the creation of filter bubbles (see chapter 2.4.) disinformation is more likely to end-up on the timeline of a person whose values and interests are similar to the content of the manipulated post. Second, the retraction of a false post hardly ever reaches as many users as the targeted fake post. For instance, in 2013, MP Magdalena Tasheva from the Nationalist Party “Ataka” claimed in the Bulgarian parliament that the public cost of one refugee is 1,100 Bulgarian lev per month, in comparison to a monthly 150 lev for pensionaries. In an article, the newspaper Capital debunked this statement as a myth. However, their article reached only 20.500 reads (Lestarska, 2013). So not every person who was misled by this disinformation was informed about its retraction. Third, as we will explain, the distribution of disinformation can be rather lucrative, which makes it an attractive source of income. Fourth, disinformation can be very powerful, because it perfectly interacts with the emotional and psychological mechanisms of human behaviour and reasoning (Brodnig, 2017, 111).

In the following chapter, we discuss some of these phenomena and mechanisms. Gaining a better understanding of why false information is so resilient is the first step to fight back against the influence of disinformation.

3.1. Politically Motivated Reasoning

Politically motivated reasoning addresses the question of how we process information we are exposed to (perceiving, evaluating, judging, reasoning, and remembering). Kraft, Lodge & Taber (2015) argue that our reasoning tends to be motivated by our political beliefs and guided by the confirmation bias. This means we tend to trust information that supports our belief system and agrees with our political and cultural values easier than opposing information.



Algorithms can enhance this effect, which may lead to the creation of filter bubbles. Because we respond (post, like, comment) more frequently to posts on Instagram or YouTube which agree with our interests and beliefs, algorithms used by social media platforms will keep showing each of us similar content again and again to keep us engaged as long as possible. As social media platforms make money by selling ads, the longer one keeps watching videos on the site, the more money the company makes.

Thus, opposing opinions or topics outside of my range of interests become excluded. Yet, functioning democracies rely on broad and diverse discussions and the exchange of pluralistic worldviews. Therefore, disinformation expert Ingrid Brodnig suggests a way out of the maze. She calls for algorithms sensitized and programmed to allow pluralistic and democratic debates. For instance, a “surprise me” button, which shows users everyday posts shared outside their filter bubble (Brodnig, 2018, 180).

3.2. The Bandwagon Effect

Politically motivated reasoning can also support radicalization when the so-called bandwagon effect comes into play. This psychological phenomenon describes how we are affected by the people around us. The more people I know who believe a specific thing, the more I will adopt their thinking. For instance, social media allows like-minded people to form homogenous groups. This can be empowering, when e.g. otherwise isolated LGBTI+ youth can connect across borders, create a safe space and openly share their thoughts, emotions, and experiences. Yet, it can create challenges for democracies and human rights, when right-wing extremists, conspiracy theorists, and religious radicals build likeminded exclusive groups in which they can radicalize their worldviews. Here, positions on specific policies that threaten human rights such as anti-gay rights become a marker of membership within these identity-defining affinity groups (Kahan, 2016).

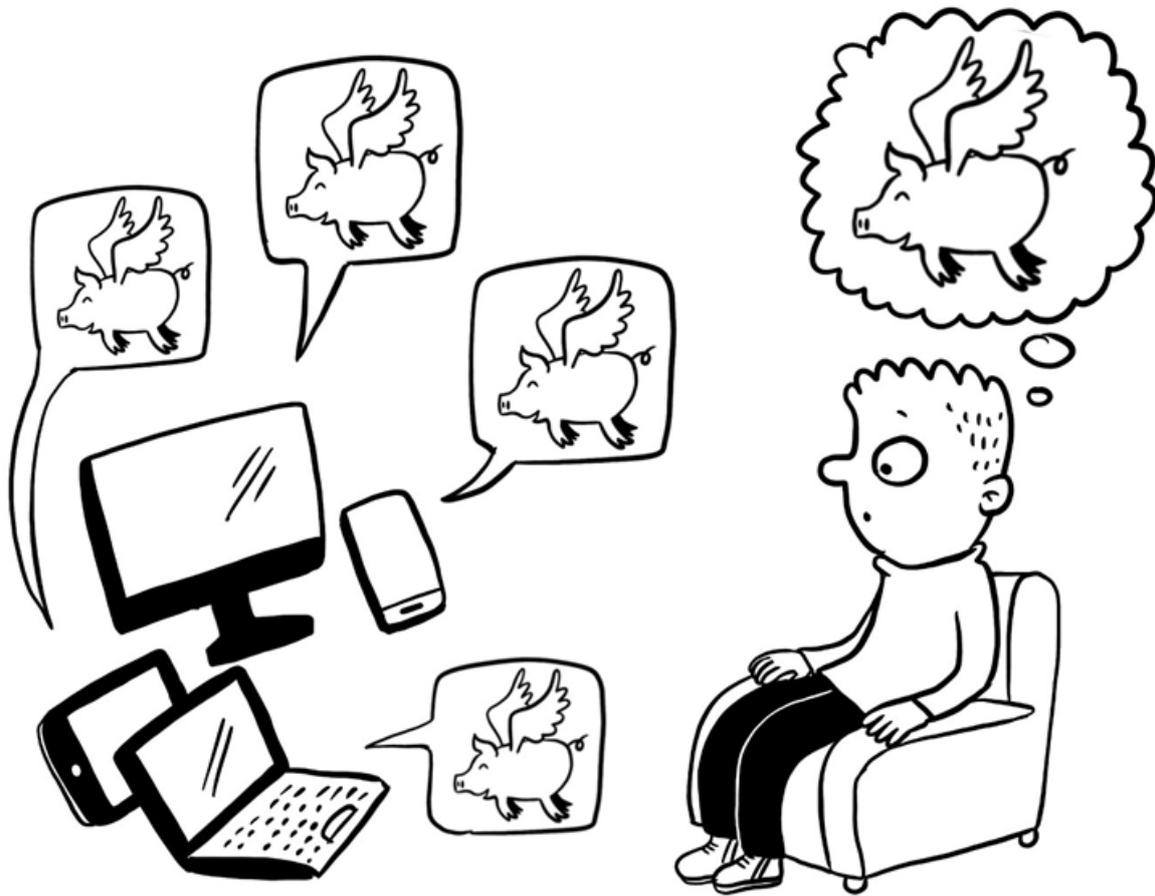
3.3. The Mere-Exposure-Effect

In 1986 late professor of Social Psychology at Stanford, Robert Zajonc, conducted research on how people make sense and navigate through social worlds. Through his studies on influence, he concluded that familiarization plays an essential role. When we are – again and again – exposed to specific information, symbols, or pictures, our affection for them increases. This is what Zajonc describes as the “mere-exposure effect” (Zajonc, 1986). This effect is used in advertisements but right-wing groups make perfect use of this effect as well when creating and spreading disinformation in two ways: first, by using new technologies such as algorithms, troll farms, and social bots that help increase the rate of interactions per post second, by focussing on a few topics which they repeatedly disseminate.

Deriving from research conducted by Ruth Wodak, professor at the Lancaster University, and expert in the field of right-wing populist discourse, disseminated topics may vary in specific national and situational contexts. However, all of them share two components. First, they stir-up hate against specific groups (i.e. Jews, Roma, homosexuals, feminists, refugees), characterizing them as threats for “us”, “our nation” or “our values”. By doing so they construct two seemingly homogenous and opposing groups: “us” vs. “them”.

Second, they provoke mistrust against established democratic institutions and promote anti-intellectualism while calling upon “common sense” and providing “alternative sources” of information. Thus, through the power of repetition and the mechanism of automatization affection can increase for even the most bizarre false information. Therefore, through a closer examination of the mass of posts created and distributed by populists online, the repetition of the same or very similar content is made visible.

Here is the good news: Author Ingrid Brodnig (2018) argues that the mere-exposure effect can likewise be used to dismantle the power of disinformation. On the one hand, understanding these psychological effects is the first step against disinformation. On the other hand, the mere-exposure effect can be likewise used as a tool against false information by repeatedly posting facts and the retraction of disinformation (118).



© Ka Schmitz

Mere-exposure Effect:

Being exposed to the same message again and again, until you believe it.

3.4. The Continued-Influence-Effect

The “continued-influence effect” refers to the continued influence of disinformation after it has been retracted. A study by Ecker, Joshua & Lewandowsky (2017) suggests that critical information of fake news “almost always continues to be used to a significant extent” (4) even after it has been corrected.

A case from the anti-vaccination debate exemplifies this phenomenon. In 1998, Dr. Andrew Wakefield published a paper in which he described a connection between autism and vaccines, drawing from his own research data. These research findings led to an increasing distrust in vaccines amongst the population, which resulted in a health crisis in Europe and the USA. Yet, this connection between autism and vaccines and Dr. Wakefield's entire study proved to be wrong. Following research showed that Dr. Wakefield's data were fraudulent. He was found guilty of scientific, ethical, and medical misconduct. However, due to the continued-influence effect of disinformation, many people still believe in the connection between autism and vaccines. Because of its impact, it is considered to be one of the most damaging medical hoaxes of the last 100 years (Flaherty, 2011). Furthermore, this hoax is still shared repeatedly in anti-vaxxer groups on social media (Wong, 2019).

Despite the continued influence effect, the authors Ecker, Joshua & Lewandowsky (2017) highlight the importance of retracting fake news. The influence of disinformation cannot be dismantled at all if it is not retracted. Hence, once disinformation is out there, its retraction must be likewise re-posted again and again to decrease the damage it will do (2).

3.5. The Problem with Pictures

Popular social media platforms like Instagram and YouTube are constructed around photo- and video-based communication. According to contemporary psychology research, images garner more attention, generate more emotive responses, and are more memorable than traditional written communication (Muñoz & Towner, 2017). Our brain simply processes pictures faster and easier than verbalised information. This phenomenon is called "the picture superiority effect" (Paivio, Rogers & Smythe, 1968). The picture superiority effect further explains why image-based platforms such as Instagram and Youtube are so popular. Consequently, Instagram and Youtube are important investment platforms for big companies to sell their products through ads, affiliate marketing or product placement.

In computational propaganda, manipulated pictures and videos, memes, and GIFs play an essential role. For example, pictures and graphs can be taken out of their context. For instance, in 2015 MP Christoph Mörgeli from the right-wing party SVP (Swiss People's Party) posted a picture showing a mass of people on a large ship and many more human beings waiting to board it. The picture was mockingly titled "the qualified employees are coming". The same picture was posted by the German ultranationalist party NPD (National Democratic Party of Germany), likewise, to spread hate against and fear of refugees. The

picture re-appeared again and again in different contexts on social media, always claiming to show African refugees trying to enter Europe today. These claims proved to be wrong. This picture was taken in 1991 after the fall of communism in Albania, showing Albanians arriving on the ship Vlora in Bari (Italy) (Neue Zürcher Zeitung 2015). Here, a 30-year old picture was taken out of its original context to stir up hate against refugees. The following case exemplifies the power of pictures as well. Additionally, it highlights the effects of disinformation on people's lives offline.

In February of 2017, a photograph of politician from the Social Democratic Party and then-Deputy Speaker of the Croatian Parliament Milanka Opačić was published on a widely read right-wing portal. The photograph displays Opačić in a red T-shirt with the Serbian national symbol of four letters "C", which was an obvious reference to Opačić's own nationality and/or political affiliation. Its purpose was to inflame readers of the portal. The article claimed that the photo was authentic and not manipulated. Several days later, other media discovered the source of the photograph, proving its manipulation. The politician's face had been attached to a photo of a different person. The police have expressed suspicions that the dissemination of the photograph was "motivated by hatred and intolerance". Opačić was given temporary police protection, as she was assessed as being at risk of assault (Hina, 2017).

3.6. When Emotions Meet Algorithms

In the previous sections, we discussed various components fake content depends on - a group or person creating the content, algorithms, social bots, etc. Yet, it comes down to another powerful agent who ensures its survival. The many people who keep liking, commenting and sharing disinformation - us. But what motivates us to engage with fake news in the first place? To answer this question, we take a closer look at emotions. Typically, disinformation is designed to evoke powerful emotions within the users, such as fear or rage. Linguist Ruth Wodak exemplified in her book *Politics of Fear* (2016), how right-wing politicians became experts in spreading fear-based messages. In her analysis of various right-wing parties across Europe, she identified how political messages are designed to evoke fear by portraying marginalised groups as threats and to legitimize the dismantling of democratic institutions. For

instance, in 2019, the minister of internal affairs of Austria, Herbert Kickl, portrayed refugees as threats to the Austrian society, while simultaneously questioning the EU Charter of Fundamental Rights and the constitutional democracy by arguing “the law has to follow politics and not politics the law” (Der Standard, 2019).

Likewise, author Ingrid Brodnig (2018) discusses this phenomenon and how disinforming content on social media is created to be spreadable. In her book on false information and technical manipulation, she illustrates how interactions (likes, shares, comments) increase when they appeal to our emotions. Especially rage, says Brodnig, is a powerful motivator for action. This is why disinformation is often designed in a sensational and polarizing style. On the other hand, YouTube and Facebook’s algorithms are coded to increase your interaction on posts and the time you spend on the platform. As the company makes money by selling advertisements, more time spent on the site guarantees a higher profit. Hence, posts generating rage or fear will be liked or shared more often. The algorithm prioritizes them in comparison to others (44). Because fake news appeals to our emotions, they become what Joshua Green and Henry Jenkins (2011) define as spreadable. Through this process of grass-roots action, by being repeatedly shared by us, disinformation becomes viral (116).



© Ka Schmitz

Disinformation is often designed to trigger rage or fear.

3.7. *It's all About the Money*

The dissemination of computational propaganda and our online behaviour, in general, helps different actors to make a lot of money. On the top of the list are social media companies (Facebook, Reddit, Twitter, etc.), marketing companies, and big brands, which use these platforms to sell us their products. Furthermore, it has become increasingly easy for individual users such as vloggers and influencers to make money if they create posts with a high clickability. For instance, during the 2016 US election, a group of Macedonian teenagers in a town named Veles made a lot of money by creating disinformation about presidential candidate Hillary Clinton and by using Google Ads. Google Ads is an advertising service allowing you to make money by advertising brands on your website. The more often someone clicks on your site, the higher your profit. According to the magazine *Wired*, one of these teenagers with the pseudonym of Boris, an 18-year-old boy, earned \$16.000 off his pro-Trump websites between August and November. He created pro-Trump posts because they showed a higher clickability. In Macedonia, the average monthly salary is \$371 (Subramanian, 2017).



© Ka Schmitz

The more spreadable content is, the higher the profit.

How to Detect Fake News

Chapter Four

Congrats! The first step is already done. To resist the power of disinformation it is essential to be aware of its existence and to educate yourself on the issue. In this chapter, we outline what you can do if your gut tells you that specific content might be falsified or even faked, or you believe a source not to be credible. We define credibility as a source of high quality, which is based on facts not false information and is, therefore, trustworthy. Identifying a source as credible or not is like solving a case. Detectives collect various clues until they make an educated decision on who committed the crime. Likewise, through questioning and using analytical technical tools (see also appendix) each user collects clues in order to identify a source as fact or fake. For further tips on detecting disinformation please see the teacher & student manuals.

4.1. Investigating the Content Creator

Credible sources always name authorship, sometimes credentials and the affiliation to other newspapers or institutions are provided as well (Schudson, 2017). There is no further information about the author available? Dig a little deeper and investigate their digital footprint to get more background information. The following questions can help you do this:

- Did the author write more about this topic in the past? Are they an expert? Look up other topics the author wrote about. Does this person have a LinkedIn page or a CV online where you can learn more about their credentials and experience?

- Do they have profiles on social media? This simple search can provide you with useful information. For instance, a person claims to have witnessed a certain event? Check out their Twitter or Instagram accounts to see if it was even possible for them to be there.
- How does their network look like? With whom are they affiliated? Check out the different organizations they are affiliated with.

4.2. Investigating the Machine : How to Identify a Social Bot

You do not find anything about this person, just a Facebook account with only one profile picture? Take a closer look. A blank page, no friends, but a high frequency of comments or tweets can be an indicator that you stumbled across a troll or a social bot. Copy-paste the picture and put it in a reverse-image search to find out if this picture was stolen. Trolls often steal someone else's profile to create the idea of being a real person.

Likewise, social bots are programmed to create fake accounts by automatically searching the internet for pictures, names, and texts, in order to act like a human user. The first bots were easier to identify as they produced a high number of posts very quickly. Nowadays, they are programmed to imitate human behaviour such as sleeping time or small talk or thinking pauses while writing a response to comments. Social media sites use captchas, tests to identify bots, which are activated when a "user" interacts in an abnormally high frequency. However, the following list of questions can help you to further investigate (Bundeszentrale für Politische Bildung, 2017):

- How many friends does this user have? Bots tend to follow a lot of users, but they have only a few to none.
- Are there any pictures on the account? Are there hints that this is a real person?
- What content is posted on this account? Is there a pattern? Bots are programmed to share/comment/like the same content repeatedly.
- What language is used? Bots only have a small range of vocabulary and use the same phrases repeatedly.

- How does the account behave? Are 30 or more posts shared every day? Humans wouldn't share that many posts in such a short period of time. When is the account active? Are there any natural patterns like time-off social media to work, sleep, do something else?
- How does the account interact? How fast does the account react to other posts? How many conversations does the account have at once?

4.3. Investigating a Website

You stumbled across an article posted by a blog/newspaper/website you do not know? Here are some things to look out for:

- Where was the content originally posted? What organization or person is responsible for this account or website? Who is contributing to this page? Who is affiliated with it? You should always know where content was first posted. Credible sources provide information on the page's objectives, involved sponsors and/or organizations, and are transparent about their finances. Check out the "About us" page. Additionally, European websites are obliged to have site notes in which important basic information is provided.
- Where is the domain registered? Is it situated in Belgium, Russia or the USA? A website's domain can give you important information about the owner's location.

4.4. Investigating a Text

Different formats (text, video, picture) demand different tools of investigation. When you want to further investigate a type of text (tweet, article, blog, etc.) to prove its credibility, these questions can be helpful (Schudson, 2017):

- What piece of text is this? Is this an opinion? Is this a parody? Is this an article? Credible sources are always open about the format they create.

- How is the spelling? What language is used? Is the language extreme, violent, does it trigger rage or is it written in a neutral manner? Credible sources attempt to collect all the facts and discuss them neutrally. Established media should have a fact-checking team that assists journalists during their work. Before an article gets published, peers review it.
- What facts are omitted? Are the sources legitimate and documented? Was the site willing to retract, correct and apologize misstatements in the past? Credible sources document and provide information on their sources. Sometimes mistakes happen. Credible sources are open about misstatements and make apologies if necessary. They do not lie about or hide them.
- Does the content creator reference other videos, interviews, articles? Are these references diverse or do they share a common emotional language, a specific point of view on social issues? Did the author cite other sources correctly or was the content manipulated/misinterpreted to strengthen their own argument?
- When was this article first published? Is this the latest news or an old video re-posted? Credible sources are transparent about when specific content was created.
- Does the author portray different points of view? Are arguments portrayed in their complexities or oversimplified in short statements? Credible sources portray different perspectives and discuss contrary pieces of evidence in their articles.
- What is the purpose of this article? Was it created to advertise a product? To inform me about a political issue or to make me laugh? To stir-up hate against e.g. homosexuals? Who benefits when you read this?

4.5. Investigating a Picture

Some of the questions listed in the previous chapter “Investigating a Text” can be likewise useful to investigate a picture (or video). For instance, what is the purpose of this picture? Is this video created to make me angry/sad/happy? However, there are some additional technical tools, which will help you verify a picture's (or video's) credibility:

- Reverse Image Search: When you verify images, it is important to work with the original data. The first tool you should use is the reverse image search. For this, you can turn to different sites like Google reverse image search or TinEye (see appendix). It is an easy way to find out when a picture was initially posted. For instance, does a picture claim to be showing an attack in Sofia. Yet, a reverse image search shows that the same picture was posted two years ago in Rome. The tool of the reverse image search will also show you similar images. This will help you to find out whether a picture was manipulated. Use different reverse image search tools to find more results.
- Geolocation: When in doubt about pictures, it can also be helpful to use geolocation tools. For instance, an image pretends to show a riot at the main square in Germany? Use the 3D mode on Google Maps and see what the main square looks like. Does it resemble the posted image?

4.6. *Investigating a Video*

Videos can be easily manipulated. Live stream videos are (so far) an exception as they are very hard to fake. When a video seems unbelievable, you should trust your instinct and further investigate. As with pictures and texts, the first thing you should do is find the user who posted the original video. Unfortunately, there is no reverse image search for videos. However, this tool can be used here as well. For instance, you can take screenshots of the first image of a video or important scenes and run them through reverse image searching tools. Sometimes the same video is cut into pieces and re-posted again. Thus, at first glance, it may seem that more people witnessed one event. Here are some additional tools for your own investigation:

- Google Translate: The video was posted in a different language? Just use Google Translate and find out what the description says.
- Unique Identifier: Some videos are re-posted and taken out of their original context. Therefore, it is necessary to know when they were uploaded. On Instagram, the time and date are embedded. Click on the three little dots next to the posted image and on “embedded”. Now copy the link and paste it into a word document. At the end of the link, you will find date and time embedded. On Twitter and Facebook, you will find the date and time next to the post.

- Amnesty International Data Viewer: Was the video posted on YouTube? For this, we suggest using Amnesty International YouTube Data Viewer to verify the date and time of a video. It automatically generates thumbnails as well which you can run through a reverse image search. Be careful, not every network automatically shows you your time zone. Twitter does if you are logged in with your own account. Facebook shows you the time selected on your computer.
- Watch frame by frame: For videos that seem “unbelievable” it is also helpful to look at them frame by frame. For this, use the tool “Watch frame by frame”. It allows you to click through a video frame by frame. When using this tool look at the different images carefully. Do shadows change or appear where they are not supposed to? Do objects get blurry or do parts of them disappear suddenly? All these clues can hint to a manipulation.

Lastly, before you start your own investigation, be lazy and use the crowd! There are many fact-checkers out there, who want to make the internet a more trustworthy place. Just google the name of the article or video with the words “fake” or “hoax”. There is a good chance that someone else did already did the research.

*Wikipedia –
Treat
with Caution!*

Chapter Five

Wikipedia is great to provide basic information on a topic. Your research journey can begin on Wikipedia, but it should never stop there. Never use Wikipedia as your single source. Wikipedia is not a reliable source. Even the creators of Wikipedia do not consider it a reliable source (Wikipedia, 2019b). Articles are written collaboratively, and authors can contribute anonymously. This is both the site's advantage and disadvantage. On the one hand, the grass-roots nature of Wikipedia allows everyone to be an author or editor. Thus, knowledge becomes democratized and information about almost every topic is collected and made accessible.

As everyone can contribute, there is always some danger of misinformation. Over the years, the page has put in a lot of effort to create a set of regulations and mechanisms to improve the quality of the articles and avoid the distribution of misinformation. Yet, you should always use the site with caution. It is important to keep the selection and content disparity between the different Wikipedias in mind. There is a great number of articles available in English, around 5,860,000 in 2019. Compared to this, there were only 2,301,800 in German, 252,200 in Bulgarian, and 205,328 in Croatian (Wikipedia 2019a). Just by looking at these numbers, you can identify an English language privilege. Yet the fact that Wikipedia is available in various languages can be a helpful tool when you are researching specific content. By comparing your research topic on different Wikipedias, you will be able to identify nuances, perspectives, and how topics are framed differently.

Wikipedia is not as egalitarian as it might appear, so it is important to analyse biases. For instance, the overwhelming majority of contributors are male (and white), which results in a systematic gender bias of the encyclopaedia in content coverage (ibid.).



Also, when you cite Wikipedia in your own work, it is important to put the date and the exact page as articles are always re-edited so your readers or listeners will not have the same version of the article you used.

Use multiple and independent sources. Only by doing so will you gain insight into the complexity of topics, make multiple perspectives visible, and will be, thus, able to form your own educated opinion on ideas and issues.

Some suggestions for using Wikipedia include using Wikipedia for basic information and collecting keywords. Wikipedia provides a reference list at the end of the article. These references can be a useful tool to continue your investigation. Recheck the sources stated in the articles. You may also use hyperlinks to get an idea about other themes your topic of choice is embedded in or connected with. Check out the editing history. This can give you cues on major discussion or perspectives within a field. Also, when you speak more than one language, you should always compare the different Wikipedias to learn more about how a topic is framed differently by the editors, what is mentioned and what is left out.

*My Digital Self
or
How to Be
a Conscious User*

*Chapter
Six*

Trolls steal profile pictures to make their account seem more trustworthy. Therefore, it is important to protect your own digital self. Under “settings” and “privacy” on your social media accounts, you can select who can see what you posted. We do not suggest public profiles. You should also protect your profile picture from abuse. For this, you can choose the option to hide your profile picture in search engines like Google and make them only visible on your social media. In general, we advise making all other social media accounts private. Private setting on social media keeps changing. Yet we highly recommend keeping your privacy in mind when using social media. If your profile has been stolen, you have to report the abuse. Contact YouTube/Instagram/Facebook/Snap Chat/reddit/8chan etc. directly. You can do this by clicking on “?” and “support inbox” or “report a problem” (Wannenmacher, 2017). To avoid filter bubbles, you should change the default settings on your research browser. You can change this under “settings”. If you need further help, check out the instructions provided in the support section of your browser. The Digital Methods Initiative (2015) created a short tutorial for Firefox.

Get active!

Do not ignore discrimination (racism, sexism, homophobia), but argue against them.

Mistrust pictures!

Our brain processes images easier than verbal information.

Do not share posts from sites you do not know!

Follow hoax detectors!

Be the first one to learn about hoaxes and disinformation by following fact-checker portals.

Have fun!

Yes, there are trolls, social bots, and haters out there, but the internet is also a place full of trustworthy information and cat content. So be a conscious user and have fun!

Educate yourself!

Be supportive!

Rage and fear are powerful emotions, but so is joy. Like, share, and comment on credible news.

Check your emotions!

Disinformation is designed to trigger hate or fear. Take a couple of breaths and think about how this post makes you feel. Become an emotional sceptic!

Talk to your friends

Discuss the spread of disinformation and hateful online behaviour with your friends.

Follow established media and journalists

Specific groups want to spread mistrust against established media. Although there is a lot of disinformation on the internet, not everything is false information. Support, follow and get your news from a diverse set of established media.

Develop a radar!

The more you think and talk about this topic, the easier it becomes for you to detect it.

Do not feed the troll!

Trolls want to spread mistrust, annoy others, and ruin the party. Ignore them, report them and support the targeted person.

Report fake news!

When you stumble across disinformation report it, spread the word, contact fact-checkers and help the retraction to get viral.



Bild © Ka Schmitz
How to Make Social Media a Safer Place



Appendix

Tools

Amnesty International YouTube Data Viewer

This tool allows you to extract hidden data from videos posted on YouTube. You can find out the exact upload time of the video, which is useful to determine which version is the original when confronted with several copies of the same video. The tool also extracts thumbnails from the video, which permits you to conduct a reverse image search and, therefore, to find an older version of the same video:

<https://citizenevidence.amnestyusa.org/>

Breaking News Generator

A website on which you can create fake news and publish them. It can be a helpful tool in order to get a better understanding of how easy it is to generate fake news:

<https://breakyourownnews.com/>

Headline Generator

Like the Breaking News Generator you can create fake news headlines with this tool. For this, you can choose between different designs simulating established media such as The Guardian, Fox News, Le Monde:

https://www.classtools.net/headline_generator/

Fake News Detector

A Chrome extension that warns users by marking fake news in red and in orange news that is likely to be fake or the links that are likely to be clickbait:

<https://chrome.google.com/webstore/detail/fake-news-detector/aebaikmeedenaijgjcfmndfknoobahep?hl=de>

Fakey

An online game everyone can play for free. It simulates a social media feed with different news and the player has to decide whether he/she wants to share, like or fact-check the post. After the player chooses an action, he/she learns whether the article comes from a reliable source or not and, therefore, if the action chosen was appropriate. The aim of this game is to learn to recognize fake news on one's social feed:

<https://fakey.iuni.iu.edu/>

Google Image

To make a reverse image search and to find out when and where a picture was originally posted. It often happens on the internet that pictures are posted to illustrate facts or news. The problem is that the pictures used can be taken from another context and used to illustrate events to which they are not linked. This practice is used in order to mislead the readers. By doing a Google image reverse search it is possible to find out where and when the picture was originally posted and to determine if it was intentionally taken out of its original context in order to trigger a specific reaction from the user.

<https://www.google.com/imghp?hl=EN>

Google Maps

To verify geolocations, measure distance, or use a 3D view to find specific buildings:

<https://www.google.com/maps>

Google Translate

To find out what a video's or picture's description means:

<https://translate.google.com/>

TinEye

TinEye is a tool that allows the user to make a reverse image search. This makes it possible to find out when and where a picture was originally posted as well as to see whether it was modified or not: <https://www.tineye.com/>

Watch Frame by Frame

To slow videos down and watch them frame by frame. This will help you to identify manipulation easier: <http://www.watchframebyframe.com/>

Waybackmachine | Internet Archive

The Internet Archive is building a digital library of Internet sites and other cultural artefacts in digital form. The platform provides free access to everyone with the aim of providing Universal Access to All Knowledge: <http://www.wayback.com/>

Wikimapia

Wikimapia is a multilingual open-content collaborative map, where anyone can create place tags and share their knowledge. It can be used as a category-based search engine (universities, shops, churches, etc.) and to verify geolocation: <http://wikimapia.org/>

Who tweeted it first?

To find out who created the first tweet on a topic: <http://ctrlq.org/first/>

Yandex

To verify geolocation especially in Eastern Europe: <https://yandex.com/maps/>

Initiatives & Laws

Bulgarisches Council of Electronic Media

The website of the Bulgarian Council of Electronic Media (СЪВЕТЪТ ЗА ЕЛЕКТРОННИ МЕДИИ). It is the official organisation on the Bulgarian national level. Its responsibilities are making decisions in connection with cases about different media, journalists, and broadcasts. Bulgarian and European Laws, as well as bad and good practices, are published on the site. Languages: Bulgarian: <https://www.cem.bg/>

The Cyberbullying Law (Cybermobbing Gesetz)

Established on 1st January 2016 in Austria. Cyberbullying is defined as "continued harassment by means of telecommunications or computer systems". This means that someone uses either telecommunication or computer systems (SMS, phone calls, emails, social media etc.) in order to unacceptably prejudice someone's lifestyle. It includes actions that harm someone's honour in front of many people as well as the action of revealing to a large number of people facts and images of the most personal sphere of a person's life without their consent. The penalty for practicing cyberbullying varies between 720 day fine and one year of imprisonment. If the cyberbullying is followed by the victim's suicide attempt or their suicide, the author of the bullying faces up to three years of imprisonment.

https://www.oesterreich.gv.at/themen/bildung_und_neue_medien/internet_und_handy_sicher_dur ch_die_digital_e_welt/3/3/Seite.1720229.html

DostaJeMrznje

The website DostaJeMrznje.org ('EnoughWithTheHatred.org') is dedicated to reporting hate speech and discriminatory speech in the public space, including the media, social networks, physical public spaces, etc. Each report is processed by a team of administrators and treated accordingly, with respect to the relevant legal provisions.

Languages: Croatian. <http://www.dostajemrznje.org/>

EU General Data Protection Regulation (GDPR)

The EU General Data Protection Regulation is a regulation in EU law on data protection and privacy that applies to all EU citizens as well as to all citizens from the European Economic Area (EEA). The aim of this regulation is to strengthen and unify the protection of personal data within the EU. It aims at giving people more control over their data, it forces companies to be more transparent on their use of personal data and to fine them when abusing data privacy, and to simplify the regulatory environment for international businesses: <https://eugdpr.org/>

Faktograf

A Croatian fact-checking portal published by the organization GONG and supported by the European Union. The portal publishes fact-checks of statements made by politicians and other relevant stakeholders in the public space and longer analytical pieces. The portal's output is almost exclusively in Croatian, with select, especially internationally relevant articles translated to English. The portal has developed an ongoing cooperation with N1, a regional news network. Languages: Croatian: <https://faktograf.hr/>

Jugend und Medien – Nationale Plattform zur Förderung von Medienkompetenzen

The Swiss national platform for the promotion of media skills. Its objective is to encourage children and young people to use digital media in a safe and responsible way. Languages: German, French, Italian: <https://www.jugendundmedien.ch/de.html>

Klicksafe.de

EU initiative in Germany. It is – like SaferInternet.at – also part of the EU's Safer Internet Programme and funded by the Connecting Europe Facility (CEF). Klicksafe is an awareness campaign promoting media literacy and adequate handling of the internet and new media. It addresses the challenge of enabling young users to handle the internet and new media critically while raising awareness of the problems they might encounter. In a nutshell, the work of Klicksafe aims at making people more conscious of the safe internet use for children and teenagers. Languages: English, Russian, Turkish, Arabic: <https://www.klicksafe.de/>

Kobuk.at

A media watch blog run by students of the "Multimedia-Journalismus" class from the University of Vienna. The aim of this blog is to watch the traditional media with a critical eye and question the information they provide. Languages: German: <https://www.kobuk.at/>

Medijska pismenost

A Croatian website focused on media literacy. It is published by the national electronic media regulator - the Agency for Electronic Media. It serves as a compendium of materials for teaching in the field of media literacy for various audiences but is specifically aimed at children and their parents.

Languages: Croatian: <https://www.medijskapismenost.hr/>

Mimikama

An Austrian organisation that was created in 2011 with the aim of counteracting and combating internet abuse, internet fraud and fake news. Its work focuses on social media such as Facebook, Twitter and WhatsApp. This allows the team to directly respond to the users' inquiries and to check the rumours and information it receives. Its main activity consists of debunking fake news, clarifying suspicious content and reacting to users' problems. The work of Mimikama allows for the protection of internet users against fishy and dangerous content online. For instance, it worked on debunking the drastically rising number of hoaxes in German-speaking countries that followed the waves of refugees. Languages: German: <https://www.mimikama.at/>

Netzwerkdurchsetzungsgesetz (Network Enforcement Act)

Also known as the Facebook-Gesetz (Facebook Act), a German law that was passed in 2017 in reaction to the growing amount of hate posts and punishable content on social media. The law obliges platform operators to offer an efficient and transparent procedure to deal with users' complaints. This procedure must be visible, always available and easy to use. According to the law, obviously illegal content must be deleted within 24 hours after the complaint was made. Content that is not obviously illegal has to be removed within seven days. This period of time can be extended if more time is needed for the legal examination of the content. Simultaneously, journalists expressed the law's potential in harming the freedom of the press, by e.g. overblocking content:

<https://www.gesetze-im-internet.de/netzdg/>

Saferinternet.at

An Austrian initiative supporting children, young people, parents, and teachers to use digital media safely, competently, and responsibly. This initiative was implemented by the European Union as part of the funding programme "Connecting Europe Facility" (CEF). Together with Stopline (an online reporting office against child pornography and national socialist reactivation) and 147 Rat auf Draht (a helpline for children, young people and their legal representatives), it forms the Safer Internet Centre Austria, which is the Austrian partner of Insafe, the Safer Internet Network of the EU. Saferinternet.at offers workshops and presentations for children, young people, teachers and parents all over Austria. It also produces some informative and educational material, such as brochures, videos, and folders. Languages: German: <https://www.saferinternet.at/>



The News Literacy Project

American educational, non-profit, non-partisan, independent programs that teach students how to know what to believe in the digital age. The project empowers teachers to provide students with the tools they need to become smart, active consumers of news and information as well as engaged and informed participants in the US democracy. The Sift® is the New Literacy Project weekly newsletter that delivers relevant media news and recent examples of misinformation. Languages: English: <https://newslit.org/>

References

- AllFacebook.de (2016) Der Facebook Newsfeed Algorithmus: die Faktoren für die organische Reichweite. AllFacebook.de Social Media für Unternehmen. Retrieved from: <https://allfacebook.de/pages/facebook-newsfeed-algorithmus-faktoren>
- Azhar et al (2018). The Anti-vaccination movement: a regression in modern medicine. *Cureus* 10(7). Retrieved from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6122668/>.
- Bhaskaran, Harikrishnan, Mishra, Harsh, Nair, Pradeep. Contextualizing Fake News in Post-truth Era: Journalism Education in India. *Asia Pacific Media Educator*. SAGE Publications Ltd. 2017, June 1.
- Brodnig, I. (2016, December 12). „Da stinkt was“: Wie Verschwörungstheorien entstehen. *Profil.at*. Retrieved from: <https://www.profil.at/oesterreich/wie-verschwoerungstheorien-entstehen-euronews-video-7806493>
- Brodnig, I. (2018). *Lügen im Netz. Wie Fake News, Populisten und unkontrollierte Technik uns manipulieren*. Wien: Brandstätter.
- Bucher, H. J. & Schumacher, P. (2006). The relevance of attention for selecting news content. An eye-tracking study on attention patterns in the reception of print and online media. *Communications. The European Journal of Communication Research*, 31 (3), 347-368. Retrieved from: <https://www-degruyter-com.uaccess.univie.ac.at/downloadpdf/j/comm.2006.31.issue-3/commun.2006.022/commun.2006.022.pdf>
- Bundeszentrale für Politische Bildung (2017, Juli 14). So lassen sich Social Bots enttarnen. Retrieved from: <https://www.bpb.de/252589/social-bots-enttarnen>
- Corner, John. Fake news, post-truth and media–political change. *Media, Culture and Society*, 2017, 39(7), 1100–1107.
- Der Standard (2019, January 23). Kickl stellt Menschenrechtskonvention infrage, Kritik von Ministerkollegen und Van der Bellen. *DerStandard.at*. Retrieved from: <https://derstandard.at/2000096888042/Kickl-stellt-Menschenrechtskonvention-in-Frage>
- Ecker, U. K. H., Hogan, J. L., Lewandowsky, S. (2017). Reminders and repetition of misinformation: helping or hindering its retraction? *Journal of Applied Research in Memory and Cognition*, 1-13. Retrieved from:
- First Draft (2017a). Viral video hoax: eagle snatches baby. Retrieved from: <https://firstdraftnews.org/en/education/course/verification-quick-start/1/lesson-1-eagle-baby/>
- First Draft (2017b). Toolkit walk-through with Malachy Browne. Retrieved from: <https://firstdraftnews.org/en/education/course/verification-quick-start/1/lesson-1-browser-set/>

- First Draft (2017c). Assessing Provenance. Retrieved from: <https://firstdraftnews.org/en/education/course/verification-quick-start/2/3-provenance/>
- First Draft (2017d). How to find a post's unique identification code. Retrieved from: <https://firstdraftnews.org/en/education/course/verification-quick-start/2/2-social-platforms-ids/>
- Flaherty, D. K. (2011). The vaccine-autism connection: a public health crisis caused by unethical medical practices and fraudulent science. *Annals of Pharmacotherapy*, 45(10), 1302-1304. Retrieved from: <https://doi.org/10.1345/aph.1Q318>
- Ford, H. & Wajcman, J. (2017). "Anyone can edit", not everyone does: Wikipedia's infrastructure and the gender gap. *Social Studies of Science*, 47(4), 511-527. Retrieved from: <https://doi-org.uaccess.univie.ac.at/10.1177/0306312717692172>
- Graber, R. & Lindemann, T. (2018). Neue Propaganda im Internet. Social Bots und das Prinzip sozialer Bewährtheit als Instrumente der Propaganda. *Fake News, Hashtags & Social Bots. Neue Methoden populistischer Propaganda*, edited by Klaus Sachs-Hombach and Bernd Zywiets, Springer VS, pp. 51-68.
- Green, J. & Jenkins, H. (2011). Spreadable media: How audiences create value and meaning in a networked economy. In: *The handbook of media audiences*, 109-127. Retrieved from: <https://onlinelibrary-wiley-com.uaccess.univie.ac.at/doi/pdf/10.1002/9781444340525.ch5>
- Heins, V. (2007). Critical theory and the traps of conspiracy thinking. *Philosophy & Social Criticism*, 33 (7) pp. 787-801.
- Hina (2017, February 27). Policija: Objavljanje fotomontaže Milanke Opačić u majci sa četiri "C" potaknuto je mržnjom. *Novilist.hr*. Retrieved from: <http://www.novilist.hr/Vijesti/Hrvatska/Policija-Objavljanje-fotomontaze-Milanke-Opacic-u-majci-sa-cetiri-C-potaknuto-je-mrznjom>
- Hummel, P. (2018). Fakten zu Verschwörungstheorien. *Fake oder Fakt? Wissenschaft, Wahrheit und Vertrauen*, edited by Carsten Könneker, Springer Verlag, pp. 187-195.
- Jack, C. (n.d.). *Lexicon of lies: terms of problematic information*. Data & Society Research Institute. Retrieved from: https://datasociety.net/pubs/oh/DataAndSociety_LexiconofLies.pdf
- Johnson, H. M., Seifert, C. M. (1994). Sources of the continued influence effect: When misinformation in memory affects later inferences. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 20, 1420-1436.
- Kahan, D. M. (2016). The politically motivated reasoning paradigm. *Emerging Trends in Social & Behavioral Science*, 1-24. Retrieved from: <https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118900772.etrds0417>
- Kraft, P. W., Lodge, M. & Taber, C. S. (2015). Why people „don't trust the evidence": Motivated reasoning and scientific beliefs. *The ANNALS of the American Academy of Political and Social Science*, 658 (1), 121-133. Retrieved from: <https://journals.sagepub.com/doi/10.1177/0002716214554758>
- Leistert, O. (2017). Social Bots als algorithmische Piraten und als Boten einer techno-environmentalen Handlungskraft. *Algorithmenkulturen: Über die rechnerische Konstruktion der Wirklichkeit*, edited by Robert Seyfert and Jonathan Roberge, De Gruyter, pp. 215-234.

- Lewis, P. (2018, February 2). "Fiction is outperforming reality": how YouTube's algorithm distorts truth. The Guardian. Retrieved from: <https://www.theguardian.com/technology/2018/feb/02/how-youtubes-algorithm-distorts-truth>
- Lestarska, D. (2013, October 11). Митът "Колко струва един бежанец" ("Myth" How much does a refugee cost). Capital. Retrieved from: https://www.capital.bg/politika_i_ikonomika/bulgaria/2013/10/11/2159110_mitut_kolko_struva_edin_bejanec/
- Muñoz, C. L., Towner, T. L. (2017). The image is the message: Instagram marketing and the 2016 presidential primary season. *Journal of Political Marketing* 16 (3-4), 290-318. Retrieved from: <https://www.tandfonline.com.uaccess.univie.ac.at/doi/full/10.1080/15377857.2017.1334254?scroll=top&needAccess=true>
- Neue Zürcher Zeitung (2015, September 2). Facebook zeigte Mörgeli die rote Karte. Retrieved from: <https://www.nzz.ch/facebook-zeigt-moergeli-die-rote-karte-ld.1714>
- Paivio, A., Rogers, T. B. & Smythe, P. (1968). Why are pictures easier to recall than words? *Psychonomic Science* 11 (4), 137-138. Retrieved from: <https://link-springer-com.uaccess.univie.ac.at/article/10.3758%2FBF03331011>
- Pariser, E. (2011). *The Filter Bubble. How the New Personalized Web Is Changing What We Read and How We Think*. Penguin Press.
- van Prooijen, J. & Douglas, M. (2017). Conspiracy theories as part of history: The role of societal crisis situations. *Memory Studies*, 10 (3), pp. 323-333.
- Roese, V. (2018). You won't believe how co-dependent they are Or: Media hype and the interaction of news media, social media, and the user. *From Media Hype to Twitter Storm. News Explosions and Their Impact on Issues, Crises, and Public Opinion*, edited by Peter Vastermann, Amsterdam UP, pp. 313-332.
- Schaar, P. (2017, October 9). Überwachen, Algorithmen und Selbstbestimmung. Bundeszentrale für Politische Bildung. Retrieved from: <http://www.bpb.de/lernen/digitale-bildung/medienpaedagogik/medienkompetenz-schriftenreihe/257598/ueberwachung-algorithmen-und-selbstbestimmung>
- Schönleben, D. (2017). Welche Social Bots gibt es und wie funktionieren sie? *Wired.de* Retrieved from: <https://www.wired.de/collection/tech/welche-social-bots-gibt-es-und-wie-funktionieren-sie>
- Schudson, M. (2017, February 23). Here's what non-fake news looks like. *Columbia Journalism Review*. Retrieved from: <https://www.cjr.org/analysis/fake-news-real-news-list.php>
- Stegemann, P. (2013, Oktober 28). Algorithmen sind keine Messer. Bundeszentrale für Politische Bildung. Retrieved from: <https://www.bpb.de/dialog/netzdebatte/170865/algorithmen-sind-keine-messer>
- Subramanian, S. (2017, May 21). The Macedonian Teens Who Mastered Fake News. *Inside the Macedonian Fake-News Complex*. *Wired*. Retrieved from: <https://www.wired.com/2017/02/veles-macedonia-fake-news/>
- Sulleyman, A. (2017, June 21). Brexit: Pro-leave twitter bots played 'strategic role' in EU referendum result, says Oxford University Institute, *Independent*. Retrieved from: <https://www.independent.co.uk/life-style/gadgets-and-tech/news/brexit-twitter-bots-pro-leave-eu-referendum-result-oxford-university-study-a7800786.html>

- Tandoc, E. & Wei Lim, Z. & Ling, R. (2017). Defining „Fake News“. *Digital Journalism*, 6(2), pp. 137-153. Retrieved from: <https://doi.org/10.1080/21670811.2017.1360143>
- The Digital Methods Initiative (2015, June 1). The research browser [video]. Retrieved from: <https://www.youtube.com/watch?v=bj65Xr9GkJM>
- Wannemacher, T. (2016, June 21). Dein Facebook-Profil: sicher in nur 4 Schritten. *Mimikama*. Zuerst denken – dann klicken. Retrieved from: <https://www.mimikama.at/allgemein/dein-facebook-profil-sicher-in-nur-4-schritten/>
- Wannemacher, T. (2017, July 11). Hilfe! Mein Facebook-Profil wurde geklaut. *Mimikama*. Zuerst denken – dann klicken. Retrieved from: <https://www.mimikama.at/allgemein/mein-facebook-profil-wurde-geklaut/>
- Wardle, C. (2017, June 4). Fake News – It’s complicated. Retrieved from: <https://firstdraftnews.org/fake-news-complicated/>
- Wardle, C. (2018). *Information Disorder: The Essential Glossary*. First Draft News. Retrieved from: https://firstdraftnews.org/wp-content/uploads/2018/07/infoDisorder_glossary.pdf
- Watson, I. (2010, March 14). Fake Russian invasion broadcast sparks Georgian panic. *CNN*. Retrieved from: <http://edition.cnn.com/2010/WORLD/europe/03/14/georgia.invasion.scare/>
- WerdeDigital (2016). Die Gefahren durch Social Bots. *WerdeDigital.at* 3(2). Retrieved from: <https://www.werdedigital.at/tag/social-bots/>
- Wikipedia (2019a). List of Wikipedias. Retrieved from: https://en.wikipedia.org/wiki/List_of_Wikipedias
- Wikipedia (2019b). Wikipedia: About. Retrieved from: <https://en.wikipedia.org/wiki/Wikipedia:About>
- Wodak, R. (2016). *Politik mit der Angst. Zur Wirkung rechtspopulistischer Diskurse*. Wien: Edition Konturen.
- Wolf, A. (2018, October 10). Verärgern dich diese 700 € Weihnachtsgeld für Flüchtlinge? *Mimikama*. Retrieved from: <https://www.mimikama.at/allgemein/700-e-weihnachtsgeld/>
- Wong, J. C. (2019, February 1). How facebook and youtube help spread anti-vaxxer propaganda. *The Guardian*. Retrieved from: <https://www.theguardian.com/media/2019/feb/01/facebook-youtube-anti-vaccination-misinformation-social-media>
- Woolley, S. C. & Howard, P. N. (eds.) (2019). *Computational Propaganda. Political parties, politicians, and political manipulation on social media*. New York: Oxford University Press.
- Woolley, S. (2016). Automating power: Social bot interference in global politics. *First Monday. Peer-Reviewed Journal on the Internet*, 21(4). Retrieved from: <https://firstmonday.org/article/view/6161/5300#author>
- Zajonc, R. B. & Mcguire, W. J. (editor) (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology*, 9(2), 1-27. Retrieved from: <http://dx.doi.org/10.1037/h0025848>



Notes