

2. How do fake news and hate speech affect political discussion and target persons and how can they be detected?

*Authors: Ines Beutel, Olga Kirschler and Sabrina Kokott
Academic supervisor: Robert Müller-Török and Nervin Kutlu*

DOI: 10.24989/ocg.v.342.2

Fake news can have effects, especially in an election or referendum context and so can hate speech. This chapter describes how these effects occur and how internet-based hate speech may turn into real-life physical violence or lead to other consequences in real life.

This chapter also focuses on how fake news and hate speech can be identified, especially in a Social Media context and, given the complexity that, at least to some extent, the line between freedom of expression and hate speech is difficult to identify. Fake news and hate speech may also be used to exercise - likely undue - influence in an organized manner, whether it be astroturfing⁵⁸ by lobby groups or influence exercised by both domestic and foreign governments or actors.

This influence is difficult to detect, however, there are means of semantic text and network analysis that may indicate such organized actions. The chapter will provide a survey of existing approaches and their respective applicability.

2.1. The distinction between Freedom of Expression and hate speech and fake news

Freedom of expression is a fundamental human right. On the one hand, it is indispensable in the human rights system, and on the other hand, it is crucial for the functioning of a democratic society. Because of its importance, freedom of expression has been enshrined in the Universal Declaration of Human Rights (UDHR, Article 19) [2-1] and in all major international and regional human rights treaties [2-2].

“Article 19 [UDHR]

Everyone has the right to freedom of opinion and expression; this right includes freedom to hold opinions without interference and to seek, receive and impart information and ideas through any media and regardless of frontiers.” [2-1]

In the European Convention on Human Rights (ECHR), this right is protected by Article 10 [2-2].

“ARTICLE 10 [ECHR]

Freedom of expression

1. Everyone has the right to freedom of expression. This right shall include freedom to hold opinions and to receive and impart information and ideas without interference by public authority and regardless of frontiers. This Article shall not prevent States from requiring the licensing of broadcasting, television or cinema enterprises.

⁵⁸ Astroturfing is, according to Merriam-Webster, "organized activity that is intended to create a false impression of a widespread, spontaneously arising, grassroots movement in support of or in opposition to something (such as a political policy) but that is in reality initiated and controlled by a concealed group or organization (such as a corporation)". Cf. <https://www.merriam-webster.com/dictionary/astroturfing> last accessed 09.12.2021).

2. The exercise of these freedoms, since it carries with it duties and responsibilities, may be subject to such formalities, conditions, restrictions or penalties as are prescribed by law and are necessary in a democratic society, in the interests of national security, territorial integrity or public safety, for the prevention of disorder or crime, for the protection of health or morals, for the protection of the reputation or rights of others, for preventing the disclosure of information received in confidence, or for maintaining the authority and impartiality of the judiciary.” [2-3, p. 12]

In addition to freedom of expression, Article 10 includes the freedom to receive and impart information without interference by public authorities or other restrictions. However, these are not absolute rights. Nations can restrict these rights if there are legitimate reasons to do so. Possible reasons might include national security, public health, or the protection of other rights [2-2]. However, the legal situation in the context of false or falsified information (fake news) or information containing hate messages (hate speech) is questionable.

2.1.1. Fake news

Regarding fake news, it is difficult to even distinguish whether it is fake news or not, because there is currently no overall or generally accepted definition. According to the general linguistic usage, it is news that has been deliberately spread falsely via the Internet or social networks. However, false statements can be differentiated. In a criminal prosecution, the decisive factor is what type of communication is involved. The dissemination of news can either be facts or personal opinions. The difference between these two types is that facts can be verified or falsified, hence they can be true or false. Contrary to that, expressions of personal opinions cannot be refuted. Correspondingly, they can neither be true nor false. Only if expressions of personal opinions are based on false facts, they can subsequently be ruled false. While personal opinions are protected by the right to freedom of expression, false facts and opinions based on false facts do not fall within the scope of this protection. If an author deliberately writes a false message, he potentially manipulates the reader. False information provided to the reader could lead to the reader taking the actions desired by the author. So, if an author deliberately puts false information into the world, he could even be liable to prosecution under e.g. German law [2-4, pp. 6-8].⁵⁹

2.1.2. Hate speech

However, the right to freedom of expression may be in conflict with other rights. The right to freedom of expression ends, where the protected interests of other persons are violated. To ensure this protection, Article 8 of the European Convention on Human Rights ensures the right to respect for private and family life.

“ARTICLE 8 [ECHR]

Right to respect for private and family life

1. Everyone has the right to respect for his private and family life, his home and his correspondence.
2. There shall be no interference by a public authority with the exercise of this right except such as is in accordance with the law and is necessary in a democratic society in the interests of national security, public safety or the economic well-being of the country, for the prevention of disorder or crime, for the protection of health or morals, or for the protection of the rights and freedoms of others.” [2-3, p. 11]

⁵⁹ <https://www.deutschlandfunkkultur.de/fake-news-vorsaetzliche-luegen-muessen-verboten-werden-100.html> (last accessed 05.12.2021)

Because hate speech is spread against other groups or individuals and also incites people, it is an abuse of the right to freedom of expression (Article 10 European Convention on Human Rights) and could violate Article 8. However, the use of hate speech is dealt with differently in different nations. For example, the American Bar Association considers hate speech to be legal and protected by the First Amendment, as long as it does not directly incite violence. There are many and various reasons why hate speech should be permitted [2-5, p. 1].

One reason to allow hate speech within the US constitutional system is the theory of the *marketplace of ideas*. According to this theory, all ideas, even the bad ones, should be heard to find the truth. In the free market, the truth should also compete with falseness. It is assumed that in the end, the truth will win this competition. Therefore, each individual should be able to communicate his or her opinions and ideas, so that the best among them can prevail [2-6, pp. 13-14]. Another reason to permit hate speech is the *democratic process*. In a democracy, any expression of opinion should be allowed. Citizens should also have access to all the information they need to educate themselves and make well-thought decisions, such as voting [2-6, p. 15]. Lastly, the theory of *personal liberty* can be applied. According to this theory, every person deserves the right to unrestricted expression, even if it contains hateful or fanatical statements. Expressions of opinion are a human being's freedom and essential for the development of one's potential. Hence, restricting them is a massive interference in human freedom and development [2-6, p. 16].

Unlike the USA, some countries ban hate speech, including Germany, Rwanda and Myanmar. These countries have already experienced that language can have a great effect. It has there historically happened that prejudice against a group has manipulated people and incited them to violence [2-6, p. 16]. In *Germany*, these scenarios happened during the National Socialism with the so-called "Jewish problem". Jewish people were massively attacked with hate speech. Dehumanizing terms were used by calling Jews vermin or snakes. In addition, ethnophobic statements and the attribution of unfavorable characteristics to Jews were used to incite hatred against them. By comparing Jewish people to animals and creating an "us versus them" feeling, mistreatment and violence against members of this group became commonly accepted [2-6, p. 21]. In *Rwanda* in the 1990s, Hutus, who constitute the majority of the population, spread hate speech against Tutsis. This encouraged ordinary citizens and militiamen to carry out mass killings. Between five hundred thousand and one million civilians fell victim, and the Tutsi population was reduced by 75 percent [2-6, p. 21]. Similar to Germany and Rwanda, dehumanization and the creation of an "us vs. them" feeling has led to violence against groups in *Myanmar*. In predominantly Buddhist Myanmar, Muslims belonging to the Rohingya people have been discriminated against for years. Since 2017, the Rohingya have been victims of brutal violence, including rape, murder, and arson. As a result, some seven hundred thousand Rohingya have fled since then [2-6, p. 23].

What these three examples have in common is that the leaders have convinced the citizens to take action against a particular group. The given reason for this incitement was that their own lives and livelihoods were in danger. They portrayed the group as a problem that could cause significant harm. In all cases, mass media were used to spread misinformation and hate speech. In today's world, non-official media, such as social media, also play a particularly important role. Through social media, anybody can create misinformation and spread it at a rapid speed. Because of this possible threat and the past experiences, the mentioned countries now try to prevent the repetition of these events through special laws, e.g. prosecuting those who praise the Holocaust in Germany [2-6, p. 25].

As shown, there are good reasons to legalize hate speech but also to ban it. To decide whether statements are admissible or inadmissible, the European Court of Human Rights follows two

approaches based on the European Convention on Human Rights. First, it examines whether the statement violates the fundamental values of the Convention. The abuse of rights is prohibited by Article 17 of the European Convention on Human Rights [2-7, p. 1]:

“ARTICLE 17 [ECHR]

Prohibition of abuse of rights

Nothing in this Convention may be interpreted as implying for any State, group or person any right to engage in any activity or perform any act aimed at the destruction of any of the rights and freedoms set forth herein or at their limitation to a greater extent than is provided for in the Convention.” [2-3, p. 14]

Finally, if the expression constitutes hate speech but does not restrict the fundamental values of the Convention, Article 10 (2) of the European Convention on Human Rights is invoked (see above). This paragraph is finally entitled to restrict hate speech if there is another legitimate interest [2-7, p. 1]. The right to freedom of expression and the right to respect for private and family life are indeed of equal importance, so the margin of appreciation should be equal [2-8, p. 17]. Therefore, it is difficult to decide, if a statement still falls below the protection of free expression or if it is hate speech, hence prohibited. The balance must be struck between the need to protect freedom of expression and the need to protect the individual's rights, respect within society, or public order. For the balances, the European Court of Human Rights has developed extensive case-law on hate speech and incitement to violence [2-2]. In doing so, they defined the following criteria: It has to be weighed up whether the contribution is in the public interest and how high the level of awareness of the person concerned is. Furthermore, the subject of the news report is decisive, as well as the prior conduct of the person. In addition to that, the content, form and consequences of the publication are evaluated and, if applicable, the circumstances under which photos were taken. Furthermore, the Court examines how the information was obtained and its true nature. Finally, the severity of the punishment is put into perspective [2-8, p. 17].

One example in which this balancing was applied is the case of “*von Hannover v. Germany (No. 2)*”. Here, two German newspapers published two photos showing an aristocratic family on vacation. The European Court of Human Rights ruled that these photos violated the right to privacy under Article 8 of the European Convention on Human Rights, as this information did not reflect the interest of the public. A third photo showed a prince in poor health. However, the health condition of the well-known prince is a case of public interest, so Article 8 of the European Convention on Human Rights was not violated.⁶⁰ Another example is the case of “*Axel Springer AG v. Germany*”. Here, a magazine published articles about the arrest of an actor for cocaine possession. The actor felt that his right to privacy had been violated, which is why the magazine was fined and prohibited from publishing further articles about the arrest. However, the European Court of Human Rights ruled in this case that these penalties were disproportionate and that the right to freedom of expression had been violated. The reason why freedom of expression outweighs the right to privacy is that the case involved judicial facts about a person known to the public. Also, the person was arrested in a public place, albeit for a minor crime. Even if the punishment was mild compared to the magazine, it still was disproportionate to the legitimate goal pursued.⁶¹

⁶⁰ <https://globalfreedomofexpression.columbia.edu/cases/von-hannover-v-germany-no-2/> (last accessed 10.12.2021)

⁶¹ <https://globalfreedomofexpression.columbia.edu/cases/axel-springer-ag-v-germany/> (last accessed 10.12.2021)

2.2. How fake news can be identified, especially in a social media context

Fake news are usually already structured differently than serious news. Often, fake news can already be recognized by their lurid writing style, their emotionally oriented texts and many exclamation and question marks. Most of the time, these texts are illustrated with spectacular pictures. This eye-catching packaging is designed to attract readers.⁶²

The following example shows an article from an online British daily newspaper. It reports that the man in the photo allegedly married a three-meter-long cobra. Allegedly, he would believe that his deceased wife reincarnated into her. With the modified phrase “You may hiss the bride”, the author tries to turn the attention to the article already in the headline via the ridiculous writing style. Already here the first skepticism would have to arise. If the source is examined further, the article can be identified as a hoax. When searching for the keywords, such as “man”, “cobra” or “Southeast Asia”, the true story behind the picture can be found. The man works as a fireman and specializes in catching snakes. The man had posted the photo on his Facebook page, the text is made up. In addition, the source “Daily Mail” is a British tabloid newspaper. These are gossip rags with lurid headlines that always take a dim view of the truth.⁶³

⁶² <https://www.lmz-bw.de/medien-und-bildung/jugendmedienschutz/fake-news/wie-kann-man-fake-news-erkennen/> (last accessed 20.10.2021)

⁶³ <https://www.geo.de/geolino/magazine/22568-rtkl-falschnachrichten-im-internet-so-erkennt-ihr-fake-news> (last accessed 21.12.2021)

MailOnline

Home | News | U.S. | Sport | TV&Showbiz | Australia | Femall | Health | Science | Money | Latest Headlines | The Queen | Royals | Prince Harry | Meghan Markle | World News | Covid-19 | Black Friday

You may hiss the bride! Man marries a 10ft COBRA he believes is his reincarnated girlfriend

- The 10ft cobra was spotted by the unidentified husband in South East Asia
- The heartbroken man said it bore a 'striking resemblance' to his dead lover
- They spend every day together playing board games and going to the gym
- Unidentified man is understood to have taken inspiration from Buddhism


By GARETH DAVIES FOR MAILONLINE
 PUBLISHED: 10:36 GMT, 11 November 2016 | UPDATED: 10:36 GMT, 11 November 2016

Share 9.2k shares 51 View comments

A man has married his pet snake because he believes his dead girlfriend came back to life as the cobra.

The snake was spotted by the man in South East Asia, who said it bore a 'striking resemblance' to his former lover.

He and the 10ft serpent now spend every day together watching TV, sharing romantic picnics by the lake, playing board games and going to the gym as a couple.



A man has married his pet snake because he believes his dead girlfriend came back to life as the cobra

Figure 7: Screenshot of the daily mail homepage⁶⁴

⁶⁴ <https://www.dailymail.co.uk/news/article-3926868/You-hiss-bride-Heartbroken-man-marries-PET-SNAKE-believes-dead-girlfriend-reincarnated.html> (last accessed 21.12.2021)

Another example shows a Twitter entry with a shark allegedly swimming on a highway. This occurrence should have taken place during Hurricane Harvey in August 2017. The entry is from an unknown person, which is rather unreliable. The better sources are websites of major daily newspapers. The image can be verified by inserting a screenshot, for example, in Google reverse search. The search shows that the shark always appears after hurricanes. So, it is a photomontage composed of several photos.⁶⁵



Figure 8: Twitter entry showing a shark on the freeway⁶⁶

As can be seen from the examples above, some measures can be taken for detection, to be sure that certain information is fake news on the Internet. This chapter will show how fake news can be recognized. Guidelines from various institutions (Landesmedienzentrum Baden-Württemberg, the European Union and the International Federation of Library Associations and Institutions (IFLA)) are used for this purpose.

2.2.1. Structure of the message

First, the structure of the message should be looked at. As described above, fake news often has a comical writing style.⁶⁷ Especially in the headline, fake news often uses capital letters and many exclamation marks. Sometimes the language is grammatically incorrect or doesn't fit the type of publication it claims to be [2-9, pp. 42-46]. In addition, the message is very much designed to reach the reader emotionally.⁶⁸ More information can be drawn from the formatting. To check the layout of the

⁶⁵ <https://www.geo.de/geolino/magazine/22568-rtkl-falschnachrichten-im-internet-so-erkennt-ihre-fake-news> (last accessed 21.12.2021)

⁶⁶ <https://www.bbc.com/news/blogs-trending-41084578> (last accessed 21.12.2021)

⁶⁷ <https://www.lmz-bw.de/medien-und-bildung/jugendmedienschutz/fake-news/wie-kann-man-fake-news-erkennen/> (last accessed 20.10.2021)

⁶⁸ Ibid.

page, it is necessary to check how the layout of the page is organized. The overall layout as well as the fonts, graphic elements and multimedia content should be coherent [2-9, pp. 42-46].

2.2.2. Consider the source

When a source is opened, it is worth taking a look at the page. Information can already be obtained from the URL (Uniform Resource Locator), especially from the TLD (top-level domain) extensions. The URL can be seen in the navigation bar of the browser. It forms the entire link of the particular position on a website. The URL has a uniform structure. First of all, it consists of a scheme, e.g., HTTP (Hypertext Transfer Protocol) or HTTPS (HyperText Transfer Protocol Secure). This is the protocol for transferring data. As the name suggests, HTTPS is the most secure version. The next level of the URL is the third-level domain, e.g., “www.”. Third-level domain names are not mandatory unless the user has a special requirement. Usually, only two levels are required. However, using third-level domain names can increase the clarity of domain names and make them more intuitive. In contrast, the second-level domain (SLD) is a mandatory part of the URL and shows the name of the website, e.g., “hs-ludwigsburg”. The TLD is the conclusion of the URL. The most commonly used TLD is “.com”. However, it can also contain geographical information such as “.de” for Germany [2-10, p. 6].

When investigating a website, it should be noted that fake URLs are often very similar to existing known URLs. Therefore, a close look should be taken at the URL. The SLDs should be popular on the one hand, but also trustworthy. Often a sign of Fake News is when the SLD consists of a large number of digits and hyphens. The TLDs should also be known. Established TLDs like “.com” or “.org” look more trustworthy.⁶⁹ [2-10, p. 12]

2.2.3. Author / Imprint

The next step is to check if there is an author or if there is an imprint. Here it is possible to check whether the website is private, institutional or governmental. In this context, it is crucial whether it is also an official account. In addition, information about the author should be available on the homepage. If it is a social media site, the profile can also be examined in more detail. It should be a trustworthy profile, possibly with a picture. In addition, a profile shows the interests of an author and whether he has already posted other articles on this topic. Sometimes there is a self-description of the author. This profile information is at least an indication of how trustworthy the author appears. In addition, other publications by the author on the Internet can be searched, as well as witnesses mentioned in the original article. For a reliable source, the authors should also be generally known on the Internet [2-9, pp. 42-46].

If no author or imprint is found, this probably indicates that the author should be disguised. In Germany, for example, an imprint obligation exists.⁷⁰

2.2.4. Comparison with other sources

To get an overall picture of the situation, it helps to research other sources and compare facts. Particular attention should be paid to the context and time period in which the information and images appear.⁷¹ For

⁶⁹ <https://www.sixclicks.de/blog/domain-endungen-auswahl#Wie%20ist%20eine%20Domain%20aufgebaut> (last accessed 09.11.2021)

⁷⁰ <https://www.lmz-bw.de/medien-und-bildung/jugendmedienschutz/fake-news/wie-kann-man-fake-news-erkennen/> (last accessed 20.10.2021)

⁷¹ *ibid.* (last accessed 20.10.2021)

the keywords such as names of people, places, companies, or products involved, it is necessary to check whether they are related to the real event. When the keywords are entered into a search engine, they should bring up the same event. The source may not be credible as well if the news seems too outrageous. The keywords should be found in other news from credible sources as well. If the message was originally written in another language, the original article should be consulted. Translation errors can also cause disinformation. The content can be checked with the help of own language skills or translation programs, of course with all their limitations. Sometimes a fake message can be unmasked or fact-checking information can be found by adding the term "fake" or "hoax" to the keyword. If one is still not satisfied, an expert such as university lecturers or journalists can be consulted [2-9, pp. 42-46].

2.2.5. Origin of a message

If a message is spread via social media, the originating message should be searched. This works with the help of search engines by entering parts of the message in the search engine field. In this way, contradictory statements can be compared and an overall picture of the situation can be obtained.⁷²

2.2.6. Plausible and actual information

In all cases, it is important to weigh for oneself whether the info presented makes sense and could be plausible. Because sometimes it already helps to switch on one's mind and to think one step further to expose possible untruths. During the plausibility check, it also helps to note whether the text, image, video or audio file has a creation date and whether it is up-to-date and plausible.⁷³ Bear in mind that such a date can be manipulated. When examining the date, it should be noted when exactly an event took place and whether this is correctly stated in the article. A specific date should also be present. Furthermore, it should be checked whether the chronological order of reported events is correct. If a location is specified, it can be examined whether the location of the event is correct [2-9, pp. 42-46].

If studies are cited, the original study can be checked to see if the information given is correct.⁷⁴ By inspecting the links provided, it is possible to determine whether the author refers to the original source [2-9, pp. 42-4].

2.2.7. Images, videos and audio files

Finally, information can be retrieved from photos, videos and other visual cues (including statistics and data) in news items. In the case of videos, images or other multimedia content, it must be questioned whether the visual element is reliable. Particular attention should be paid to signs of manipulation, such as filters, retouching or the like. The image could be a fake. It is necessary to pay attention to whether the medium matches the previous information. e. g. whether the date and time match the event. If there is a credit for the visual element, the authenticity can be investigated by checking the source [2-9, pp. 42-46]. Meanwhile, images, videos and audio files look deceptively real, which is why they are very difficult to identify as fake. In case of mistrust, screenshots can be entered into Google Image Search, for example. The YouTube Dataviewer⁷⁵ can also show the exact upload time of a YouTube video and preview images.⁷⁶ When using data and charts, it can be questioned whether the numbers and statistics

⁷² ibd. (last accessed 20.10.2021)

⁷³ ibd. (last accessed 20.10.2021)

⁷⁴ ibd. (last accessed 20.10.2021)

⁷⁵ <https://citizenevidence.amnestyusa.org/> (last accessed 20.01.2022)

⁷⁶ <https://www.lmz-bw.de/medien-und-bildung/jugendmedienschutz/fake-news/wie-kann-man-fake-news-erkennen/> (last accessed 20.10.2021)

are used plausibly. The figures could be manipulated. Therefore, it is necessary to check whether similar numbers for the same topic can be found elsewhere [2-9, pp. 42-46].

These are the most crucial issues to recognize fake news. But not everything false is also fake. Information can be changed intentionally (disinformation) or unintentionally (misinformation). The spreading of misinformation is therefore not fake news, but may simply be a mistake, bias or some other form of incorrect reporting. As already described in the previous chapter, a distinction must be made between a fact-based report and a *personal opinion*. In the case of an opinion, there is greater freedom due to the freedom of expression. It is therefore important to note whether the content of a medium is a personal opinion or a fact-based report. Some texts may be meant in a humorous sense, e.g., as a joke or satire. *Jokes* are short stories or exposition with a surprising twist or punch line designed to make the reader laugh. Jokes are usually easier to recognize than satire. *Satire* is more serious humor with a type of writing in which circumstances or problems are addressed in an over-exaggerated, ridiculous form. It often works with exaggerations or understatements, with ambiguities or irony. In satires, people's faults and weaknesses are pointed out, often indirectly criticizing the human condition, but mockingly and humorously. Thus, an altered piece of information may be legitimate. In this case, it is particularly worthwhile to look at the source where this message is published, because it may already be a satire page.⁷⁷ Furthermore, there are always cases of *incorrect reporting*, which also belong to the term misinformation. This happens when serious sources include manipulated content in their reporting. For example, in the case of the alleged attack in Kongsberg (Norway), a fictitious perpetrator's name was published in the media. The name was originally spread on social media by so-called trolls who deliberately wanted to confuse.⁷⁸

For a brief overview of how to check for fake news, the International Federation of Library Associations an Institution (IFLA) has created an overview that can be viewed in the figure below:

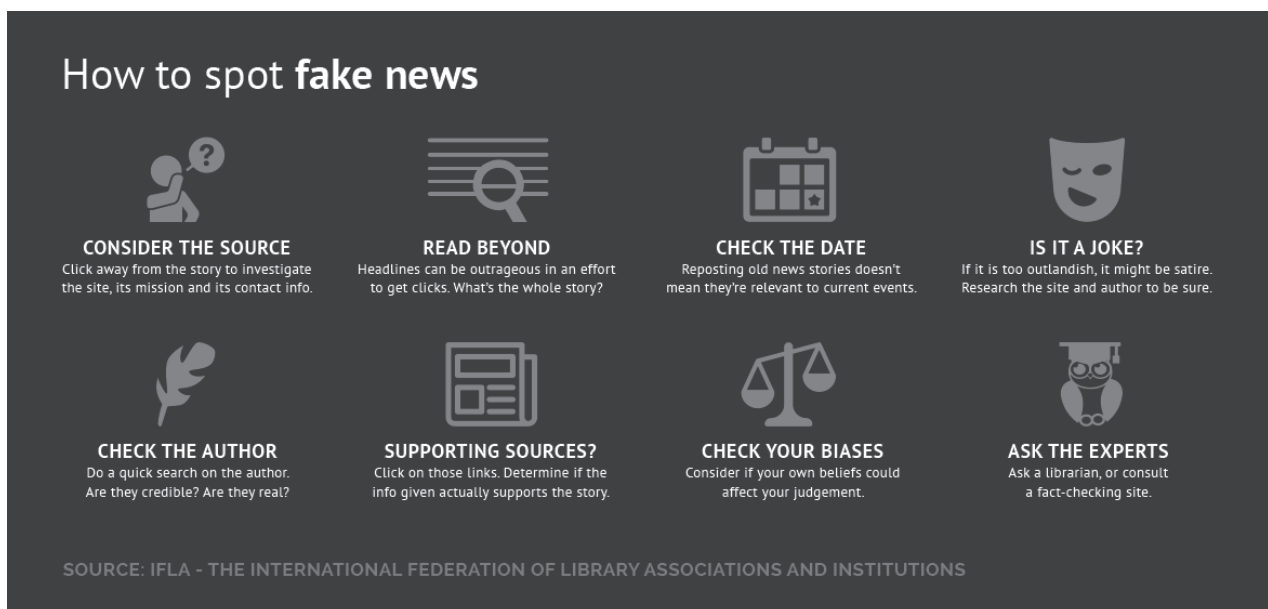


Figure 9: How to spot fake news⁷⁹

⁷⁷ <https://www.studienkreis.de/deutsch/satire-definition-merkmale/> (last accessed 17.11.2021)

⁷⁸ <https://www.tagesschau.de/faktenfinder/kongsberg-medien-101.html> (last accessed 18.11.2021)

⁷⁹ <https://www.coe.int/en/web/human-rights-channel/fake-news> (last accessed 13.10.2021)

It is important to recognize fake news. Wrong information can influence people's attitudes and actions. The next chapter, therefore, discusses the areas in which this influence can have an impact.

2.3. How can hate speech be identified? Identify the conditions conducive to the use of hate speech

Hate speech can take the form of written or spoken words, or other forms such as pictures, signs, symbols, paintings, music, plays or videos. It also embraces the use of particular conduct, such as gestures, to communicate an idea, message or opinion (cf. [2-11], p. 2). To determine whether a statement is hate speech, it is necessary to define what is meant by hate speech. Aiming to create a consistent understanding of hate speech in this book, a definition is created in Chapter 1. This delimits hate speech as follows:

“Hate speech is to be understood as the advocacy, promotion or incitement in any form of denigration, hatred or disparagement of any person or group of persons, as well as any harassment, insult, negative stereotyping, stigmatisation or threat concerning such person or group of persons, and the justification of any of the foregoing on the grounds of 'race', colour, descent, national or ethnic origin, age, disability, language, religion or belief, age, disability, language, religion or belief, sex, gender identity, sexual orientation and other personal characteristics or status, as well as the form of public denial, trivialisation, justification or approval of genocide, crimes against humanity or war crimes found by courts of law, and the glorification of persons convicted of committing such crimes.”

Put simply, hate speech involves attacking people in a discriminatory way. It must emerge that discriminatory words were used pejoratively in reference to a population group. Furthermore, hate speech has different characteristics. Laaksonen et al. defined five categories in their study. They define hate speech as messages that (1) incite violent action, (2) calls for discrimination or the promotion of discrimination; (3) attempts to degrade human dignity based on characteristics; (4) involves a threat of violence or the promotion of violent action; (5) or is accompanied by contempt, solicitation, name-calling, or slandering [2-12, p. 7].

However, it is very difficult to detect hate speech in everyday life. Only in the case of swearwords are a clear appearance of hate speech. The discriminatory must be pejoratively used in reference to a population group. But population groups can also be disparaged or denigrated without any form of expression. These manifestations of hate speech are difficult to recognize because they usually appear harmless at first. Restricting it to individual words would therefore not be helpful. Language only becomes hate speech in the context in which it is used. Especially in social media, the manifestation of hate speech is diverse and difficult to detect. Forms such as jokes, satire, or similar appear here, which are not decoded as such without context. Many people do not perform a context analysis and cannot recognize that the post is meant as a joke [2-13, pp. 338-340].

One way of approaching a medium to identify hate speech is presented below. A study by Patton et al. (2020) [2-13] is used for this purpose. They use the Contextual Analysis of social media (CASM) to detect hate speech by specializing in gang violence. The CASM works, in addition to the natural language processing tools, with the differences in geographic, cultural, age-related variance of social media use and communication. To determine the context of a post they use the following steps:

1. Baseline Interpretation

In the first step, a baseline Interpretation must be done. The message can be viewed soberly, just the context is disregarded and only the cover is interpreted. The text, emoji, hashtags, memes, images,

and videos can be used to get a first impression of whether it could be a hate message. Since these interpretations do not yet take the context into account, initial interpretations can still be influenced by prejudices, for example.

2. Examination of all biographical and offline information

To now take a look at the actual message of the message, all biographical and offline information is now to be used. First, the original social media post can be searched for specific mentions of names, communities, groups, schools, streets, local institutions or events. If the message can be assigned to a specific group, it is necessary to look for characteristics relevant to that group, such as words, phrases, emojis and other features. Contextual or cultural features can be located with the help of web-based resources. It should not be forgotten that hate speech can also be a matter of mistranslation or that other cultures have different ways of communicating. Second, the author who wrote the message is considered in more detail. All biographical information can be used for this purpose, e.g., name, date of birth, neighborhood, city. Photos can also be used to gather information on location, gang affiliation, peer network and environment. There may already be other postings where a pattern can already be identified. In this case, it can be compared whether the post matches the original postings. The final step is to take a closer look at the people tagged (@) in the post, like, share, or comment on the post. It can be asked; which relationship the persons have to the author or why these are linked. For the people who reply or comment on the post, there is also the question of the connection to the author and the reason why they are commenting on the post. Perhaps they are attracted to certain content. Possibly an intention can be discerned, namely whether they are trying to escalate or de-escalate the post [2-13, pp. 339-340].

3. Interpretation & Contextual Analysis Assessment

After the extensive context analysis, the original perception can now be reflected. Presumably, some perceptions could be downplayed or exacerbated [2-13, p. 340]. Similar to hate speech, there is a certain framework where hate speech is legitimate. In *satires*, for example, certain topics are presented with a high degree of exaggeration or ironically. This can make this particular type of humor seem like hate speech.⁸⁰ But the *time reference* is just as decisive. Over time, various terms have now been portrayed as discriminatory swear words. If older texts are used as sources, they appear as discriminatory with the time reference to today. An example of this is the work of Albert Schweitzer. From 1912 till his death in 1965, Schweitzer worked as a physician in Central Africa, founded the Hôpital Albert Schweitzer in Lambaréné (Gabon) in 1913 and was even awarded the Nobel Peace Prize for his work. Today, however, Albert Schweitzer is, by some people, judged critically because of his paternalistic attitude towards the Africans.⁸¹

2.4. Effect on political discussion, democracy, economy and society

After now knowing how to identify fake news and hate speech, it is essential to understand why such identification is important.

The role of the press in a democratic society is a vital one. The European Court of Human Rights has repeatedly underlined that the press and other media have a special role in a democratic society as the purveyor of information and public watchdog (cf. [2-14], p.6). Disinformation often highlights

⁸⁰ <https://www.studienkreis.de/deutsch/satire-definition-merkmale/> (last accessed 17.11.2021)

⁸¹ <https://www.srf.ch/news/panorama/kritische-stimmen-zum-albert-schweitzer-jubilaem>, 24.03.2013 (last accessed 20.01.2022)

differences and divisions, whether they be between supporters of different political parties, nationalities, races, ethnicities, religious groups, socio-economic classes or castes ([2-15], p. 41). Hate speech can reflect or promote the unjustified assumption that the user is in some way superior to a person or a group of persons that is or are targeted by it. This assumption might lead to certain behaviors and thoughts, in the worst case, it might even end with physical attacks. Those various effects of fake news and hate speech will be discussed in the following paragraphs and shall underline the importance of fighting them.

2.4.1. Effects in general

Before going into deep about what effects fake news and hate speech have on several sectors, the term “effect” must be defined.

In this context “effect” shall be understood as an aimed or unaimed impact on something or someone. Sometimes this impact can’t be known right from the get-go of an action. Most of the time people do not think about the outcome of their actions and how they might influence others (how easy it is to influence people and why will be discussed in chapter 2.5).

Coming back to fake news and hate speech, there are various effects and therefore impacts on other people and whole sectors.

Those impacts can be divided into three main groups:

- impacts on supporters/followers
 - fake news and hate speech might convince them of false facts
 - fake news and hate speech might activate them and encourage them to take action, ranging from postings to the use of violence
- impacts on opponents
 - fake news and hate speech might demotivate them in sticking to their opinions, continuing their role as a politician, starting a career as a politician
 - fake news and hate speech might disturb their actions and prevent them from doing what is necessary
- impacts on impartial individuals
 - fake news and hate speech might make them question their opinions and truthful news
 - fake news and hate speech might make them share the information with other people, even if they don’t believe it, hence they spread it further

Although fake news and hate speech surely are influencing other sectors too, the following chapters will deal with the effects on political discussion, democracy, economy and society. These effects will be analyzed, based on the three main groups that have been introduced in this chapter (supporters/followers, opponents and impartial individuals).

2.4.2. Effects on political discussion

Nowadays, the world is connected more than ever and slowly but steadily becoming a global village. There is a lot more communication between politicians of other countries but also between politicians in the same country. Without a doubt, the internet and social media made it a lot easier to exchange thoughts, opinions and other information.

Over time, the internet developed and became one of the main sources for people to inform themselves about what is going on in the world and therefore also became a powerful instrument that might cause huge damage when being abused. Fake news and hate speech are perfect examples of such abuse.

Because politicians and their debates on upcoming law and other rules are often very present in modern media (such as television, radio and of course the internet) and ruled as big influencers on society, they often are pulled into the spotlight and become the target of such fake news and hate speech attacks. Political discussions can cause society to split up into two sides and influence the interaction between people in real life. This could be observed during the US Presidential election campaign 2020 and the discussions between Donald Trump and Joe Biden. It appeared that America was divided into Republicans and Democrats, fighting each other and standing up for their vision of how America should continue.⁸²

When people are in such a dispute over something, fake news and hate speech can turn into a real weapon. Disinformation campaigns are the policy of “promoting lies, half-truths, and conspiracy theories in the media” [2-16, p.7]. Furthermore, a disinformation campaign can be a non-military measure for achieving political goals. The Russian Minister of Defense for example describes information as “another type of armed forces” [2-15, p.34]. Especially Russia is renowned for internet trolls, who “attack critical articles about Putin or Russian politics in European and U.S. online media, disseminate fake news [...] and distort the representation of events on heavily funded Russian export media” [2-16, p. 7]. This demonstrates how fake news are used to influence peoples’ minds and opinions on certain topics and by that achieve an advantage in political discussions.

The usage of fake news and hate speech, in particular, their spreading over social media and other news pages, can lead to severe consequences. It might lead to some politicians resigning their political function and status or in a worst-case scenario even to physical violence.

One example for that is the ‘Querdenker’ movement in Germany, where people are coming together, believing that COVID-19 is all a political setup and used by politicians to make the population bend its’ will. Such movements can be seen all over the world right now. People do not accept the measures that are introduced by governments.

One who educates himself/herself by reading truthful articles, researching current hospital figures or talking to diseased people, can easily find out that COVID-19 is definitely not fake, and thousands of people are fighting for their lives daily. On the other hand, there are a lot of theories around COVID-19 and also a lack of appreciation linked to the rules/laws and decisions made during COVID-19.

⁸² <https://newsroom.ucla.edu/magazine/2020-election-trump-biden-divided-america> (last accessed 11.12.2021)

QUERDENKEN

Figure 10: The symbol of the ‘Querdenken’ movement in Germany⁸³

2.4.2.1. Impacts on supporters/followers

The three main groups that have been introduced before, can also be transferred to the COVID-19 discussion. Some people are supporting and following those, who spread fake news about COVID-19 and believe that it’s just a political setup to restrict their rights. As the number of supporters and followers grew, so did the commitment to fight for their freedom, hence they have been activated. Activated to go on demonstrations against the political measures and also physically defend themselves. This defense is directed against policemen during demonstrations,⁸⁴ but also towards individual politicians. The main targets of these physical violent acts during the pandemic are health ministers, virologists and others who are involved in the COVID-19-debate. An example of how far those violent acts can go is a recent parade carrying lighted torches in front of the home of Saxony’s health minister.⁸⁵ Unfortunately, during the last two years, the number of online death threats against politicians who support pandemic restrictions has increased and put them under enormous emotional and physical.⁸⁶



Figure 11: Example of hate comments against the German epidemiologist Karl Lauterbach on Twitter⁸⁷

⁸³ <https://querdenken-711.de/> (last accessed 11.12.2021)

⁸⁴ <https://www.tagesschau.de/ausland/corona-protest-bruessel-103.html> (last accessed 12.12.2021)

⁸⁵ <https://www.zdf.de/nachrichten/politik/corona-protest-gewalt-verfassungsschutz-warnt-100.html> (last accessed 12.12.2021)

⁸⁶ <https://www.dw.com/en/covid-german-politicians-scientists-face-threats-online/a-56589911> (last accessed 12.12.2021)

⁸⁷ <https://www.dw.com/en/covid-german-politicians-scientists-face-threats-online/a-56589911> (last accessed 12.12.2021)

Overall, the criminal offenses against public officials are increasing year by year. So do violent offenses. In Germany for example, 1.674 criminal offenses against public officials were registered in 2019, 89 of them were declared as violent offenses. When comparing these numbers to the incidents in 2018 (which were 43), the number of violence offenses more than doubled.⁸⁸ Although this might be shocking, there are expected to be a lot more cases that have not been reported, hence the real numbers could be likely higher.

2.4.2.2. Impacts on opponents

In a German study on violence against local politicians by KOMMUNAL (a magazine on local politics) and the opinion research institute Forsa, 2.494 mayors in Germany were asked about their experiences.⁸⁹ One of the main realizations is that violence against politicians is no longer only happening in bigger local authority districts but also in small villages. Most of the violent acts are happening at public events and working offices but violent acts are also starting to affect private actions. The affected persons are complaining about being insulted, threatened, and even physically attacked.



Figure 12: Overview of the results of the German study on violence against local politicians⁹⁰

How does that affect politicians and their will to continue with or even start a political career? This question deals with the second main group, the opponents, and what impact fake news and hate speech have on them.

In this context, the study points out that the will to continue a political career or even start one decreases. Nobody wants to deal with violent acts, especially when they are carried out of the career-

⁸⁸ <https://www.bpb.de/apuz/im-dienst-der-gesellschaft-2021/329322/gewalt-gegen-amtstraeger> (last accessed 26.10.2021)

⁸⁹ <https://kommunal.de/kommunalpolitiker-umfrage-2020> (last accessed 26.10.2021)

⁹⁰ <https://kommunal.de/kommunalpolitiker-umfrage-2020> (last accessed 12.12.2021, own translation)

life into the private-life. This result proves, that fake news and especially hate speech demotivate politicians in their actions and also in their political careers. But not only that, the consequences of fake news and hate speech are going a lot further and affect the private life of politicians and their families so that they have to live in fear for their own lives and the life of their loved ones.

2.4.2.3. Impacts on impartial individuals

Finding out about the impact fake news and hate speech have on impartial individuals is not easy because they're often acting in the background, without anyone noticing it. Compared to supporters of fake news and hate speech, impartial individuals are not as outstanding and present in the media. One impact that all the fake news might have on impartial individuals is that they make them question their point of view and their opinions so that there is a potential threat, they might start to believe them someday and become followers and supporters.

On the other hand, impartial individuals are discussing a lot, because they are standing in between two opinions and try to figure out arguments for and against each side. Therefore, they might share fake news to discuss them with family and friends and that also causes them to spread further.

Nevertheless, the quick way of communication by messages and sharing links and information online within seconds makes it possible for fake news and hate speech to be spread all over the country and even beyond. Because technology and messengers are constantly being used in our daily lives and becoming more important, even essential in ways of communication, we must expect that the numbers of criminal offenses against public officials will increase further without action against that.

2.4.3. Effects on democracy

As the introductory chapter already pointed out, freedom of expression is an enjoyment required by democracy.⁹¹ Thinking of the small line between freedom of expression and hate speech, it is quite obvious that hate speech and also fake news affect democracy and the standards within. When social media was first implemented in our daily lives, everybody, including politicians, thought that social media would help to make democratic information available and help voters to make more informed choices during an election [2-17, p. 12]. What they did not have in mind is that social media can be misused and therefore affect opinion-building in a negative or simply untrustworthy manner.

One big problem for democracy was mentioned by the US political analyst Charlie Cook: “the wall between real journalism and fake journalism is becoming blurred and sometimes invisible. When people doubt the credibility of legitimate journalism, people are robbed of the facts that underlie our entire democratic process. Elections depend on citizens making informed decisions, but that’s impossible if raw sewage is polluting their news feed”.⁹²

This represents a challenge for democracy, and in particular for the electoral processes throughout Council of Europe member States, affecting the right of freedom of expression, including the right to receive information, and the right to free elections.⁹³ “While there is no doubt that in a democracy all ideas, even though shocking or disturbing, should in principle be protected [...], it is equally true that not all ideas deserve to be circulated” [2-14, p. 9].

⁹¹ <https://www.legislationline.org/documents/id/8226> (last accessed 20.10.2021)

⁹² <https://www.coe.int/en/web/human-rights-channel/fake-news> (last accessed 13.10.2021)

⁹³ <http://assembly.coe.int/nw/xml/XRef/Xref-XML2HTML-EN.asp?fileid=28598&lang=en> (last accessed 20.10.2021)

While there is fake news, on one hand, spreading untrue information about a topic, there also is news on the other hand that isn't untrue but is promoted to be fake news by politicians. One good example of that is the election campaign and usage of social media by Donald Trump. In his daily tweets, he is purposefully using wrong information to influence another person (Barack Obama and his place of birth), social group (lies about illegal immigrants), country (Mexico and its' population) or organization (World Health Organization) in a negative way [2-18, p. 13].



Figure 13: Example of a tweet made by Donald Trump⁹⁴

Whenever a news magazine posted an article that wasn't supporting his point of view or was even criticizing his way of leading, he would instantly call it "fake news media" and accuse them of trying to tear the country apart [2-18, p. 13].

Professor Tarlach McGonagle, a senior researcher at the Netherlands Network for Human Rights Research, is describing Trump's behavior as a witch hunt where accusations are made in the expectation to inflict public mistrust for media and daily press. In his opinion, this contributes to aggression and hostility towards journalism and media in general [2-18, p. 14].

The main problem is that the opinion of other people in political discussions is often portrayed as wrong, rather than different and accused of destabilizing democracy. As human beings, we are drawn to believe people with the same views and opinions on certain topics. This effect is reinforced when we are hearing a lot about those people in the media.

All societies are experiencing an increasing form of influence by journalism and media. The biggest influence is given by policy and the economy (power and money). Those, who have an excessive amount of political or economic power, are using the media to flaunt it and let it work in their favor [2-18, p. 15].

⁹⁴ <https://www.stiftung-nv.de/de/publikation/kurzanalyse-zu-trumps-crime-tweet-deutschland-viel-aufmerksamkeit-wenig-unterstuetzung> (last accessed 12.12.2021)

2.4.3.1. Impacts on supporters/followers

Transferring this cognizance to the impact fake news and hate speech have on supporters and followers, it appears that they can be highly influenced in their opinion-building on democratic votes and views. At the same time, fake news and hate speech might lead them to mistrust the rules and values a democracy is built on. Furthermore, the transparency problem with politics and laws over all is playing a big role in this case. If people could easily find truthful information about current democratic topics and easily understandable articles from the officials directly, not as many people would come across fake news and hate speech or at least would not trust them within seconds.

In contrary to the activation of supporters and followers in the context of political discussion, the activation in the context of democracy is not directed at certain politicians, but at the whole democratic system. Therefore, people are activated to question the system in terms of democratic bases and ask themselves if the politicians are still acting within the boundaries of their legitimation.



Figure 14: Symbol for a democratic election⁹⁵

2.4.3.2. Impacts on opponents

In a democracy, citizens and politicians to a certain degree are dependent on agreeing, on what is real and what is not.⁹⁶ Fake news and hate speech can destroy this consensus by offering an alternative reality. Consequently, information-based decision-making is a lot harder.

While supporters and followers of fake news and hate speech are questioning democracy, it is becoming a lot harder for politicians and also democrats, in general, to convince them otherwise. Multiple scandals, for example, the face mask scandal in Germany during the pandemic, where politicians of the CDU and CSU received commission payments for conveying purchase contracts for face masks, have shattered their credibility.⁹⁷

Because fake news and hate speech are spread extremely fast over social media and the internet in general, truthful, and most times less interesting news, are moving into the background and overpowered by disinformation.

⁹⁵ <https://www.osce.org/odihr/463626> (last accessed 12.12.2021)

⁹⁶ <https://www.dw.com/de/fake-news-sind-eine-bedrohung-f%C3%BCr-die-demokratie/a-37033453> (last accessed 13.12.2021)

⁹⁷ https://www.zeit.de/politik/deutschland/2021-03/maskenskandal-cdu-affaere-csu-jens-spahn-georg-nuesslein?utm_referrer=https%3A%2F%2Fwww.google.com%2F (last accessed 12.12.2021)

As a consequence, Democrats and politicians have to work even harder to regain the trust of all skeptics. How they could do that will be analyzed in chapter 5.

2.4.3.3. Impacts on impartial individuals

Because some true news is ruled fake news by politicians, individuals might get into believing the verdict of the respective politicians and hence believe that these are fake news. This is like the reputation effect under limited knowledge in economics [2-19, pp. 488-500]. This again might lead them to stick to one side or another. If they conclude that actual true news is fake news, the supporters and followers are growing in number. If they remain impartial, it sure will become harder for them to find the difference between fake news and facts, because those lines are blurred out more and more.

2.4.4. Effects on economy

Social media is a big business, orientated on profit and growth rather than sharing truthful information. According to the New York-based Data & Society Research Institute “the financial dependence on Facebook for content distribution has [...] weakened the reach of solid journalism. The role of journalism now is to “give people what they want” or “what matters to them” and is embedded in the same logic that drives Facebook’s algorithmic personalization and ad-targeting products”.⁹⁸



Figure 15: Illustration of multiple social media platforms⁹⁹

An interesting aspect in the context of the economy is “Disinformation as a Service” (DaaS). The difference between disinformation and misinformation was already pointed out in the introductory chapter, hence these terms will not be defined again here.

Nowadays the internet is a powerful and sometimes underestimated instrument to influence people and companies. One way to do so is by using the help of DaaS.¹⁰⁰ As the term already indicates,

⁹⁸ <https://www.coe.int/en/web/human-rights-channel/fake-news> (last accessed 13.10.2021)

⁹⁹ <https://www.som-onlinemarketing.com/2018/05/22/so-knackst-du-den-code-fuer-die-social-media-algorithmen/> (last accessed 13.12.2021)

¹⁰⁰ <https://preveny.com/desinformationskampagnen-als-gefahr-fuer-die-wirtschaft/> (last accessed 18.11.2021)

disinformation campaigns are offered as a service to spread fake and harmful news against a certain person, company, etc. These services are offered in the Darknet and pretty much everybody can take advantage of them by spending some money and defining whose reputation shall be damaged. Concrete examples for such DaaS-Campaigns are fake comments on social media, fake recommendations on company websites and their products, or spreading rumors all over the internet and in the daily press.¹⁰¹

Before those services were provided, it was quite costly to start a disinformation campaign. They were mainly implemented by criminals and unscrupulous governments underground.¹⁰² Over the years, a lot of DaaS suppliers have emerged and started advertising their “products” routinely to the private sector. Hence, it has become a serious market with lots of different offers and interested customers. What makes this dangerous is that on one hand these campaigns are inexpensive to create and distribute at scale,¹⁰³ but on the other hand can cause major damage to those people who are targeted.

The reason why it is inexpensive lies in human nature itself. A 2019 MIT study examined to which extent fake news is spread faster than the truth. The result is that “falsehoods are 70 percent more likely to be retweeted [...] than the truth” and reach the first 1.500 people six times faster.¹⁰⁴ Another interesting result of the study is the fact that this phenomenon is more pronounced for false political news rather than false news about science, terrorism, natural disasters or financial information. Somehow people are more interested in stories and scandals in the political environment than in any other environment. Finding out about something that isn’t compatible with our expectations or the way we think things should be, makes us talk about it and also share it with friends and family, especially online. As a human, we all have certain expectations and behavioral standards that we have learned and accepted throughout our lives. Whenever news is reporting anything contrary, it instantly makes us think about it, talks about it, and leads to a certain view or opinion about the person that has been involved. As a consequence, it is pretty easy and inexpensive for DaaS suppliers to satisfy their customers' wishes because the news will spread almost automatically after they have been implemented.

While fake news and hate speech are boosting one’s finances, they destroy another one. According to current studies, fake news costs the world economy more than 78 billion US dollars annually.¹⁰⁵ The reason for this loss of money is that fake news is more and more used to harm companies and their daily operations. By spreading fake news about one company over the internet, a lot of its customers might read them and question themselves, whether they still want to support the company or not. Not only are these companies losing their customers, but it also makes it a lot harder for them to gain new clients because their reputation has been damaged by the fake news spread. The consequence is a deep cut in their income. Adding upon this, they might have to put in quite a lot of money to rebuild their reputation and gain back some customers, which of course costs money and a well-thought strategy. If this strategy does not work out as planned and the reputation couldn’t be rebuilt, the companies might end up in financial ruin.

2.4.5. Effects on society

The effects of fake news and hate speech on society are closely connected with those on democracy and political discussion. People who want a radical change use the internet to spread hate speech

¹⁰¹ Ibid.

¹⁰² [Disinformation attacks in corporate sector: PwC](#) (last accessed 24.11.2021)

¹⁰³ ibd.

¹⁰⁴ [Study: False news spreads faster than the truth | MIT Sloan](#) (last accessed 24.11.2021)

¹⁰⁵ <https://preveny.com/desinformationskampagnen-als-gefahr-fuer-die-wirtschaft/> (last accessed 18.11.2021)

and fake news and by that try to convince others from their views and theories.¹⁰⁶ This is in principle what other political activists, parties, etc. also do – the difference being that they do not facilitate hate speech and fake news.

Also political and religious tension is spread more rapidly. Haters are following their goal to exclude specific groups from society and silence their voices. It is kind of contrary to what we expected from the internet: a great basis for communication all over the world and the forming of a global community based on the values of the UDHR, the US constitution and the European Convention of Human Rights. Instead, the internet is drifting in the opposite direction and separates us all into tiny groups of people who are sharing the same opinions within these groups and cutting themselves off of other people with different opinions.

This separation causes more problems that might not be visible at first sight. While one group is actively promoting hate speech about minorities, more and more people are noticing that and joining in. However, on the other side, those social minorities are experiencing severe effects and are excluded even more. Especially hate speech causes a lot of them to question their behavior and ask themselves whether there is something wrong with who they are or what they do. This results in an emotional burden and causes physical symptoms [2-20, p. 29]. Depression, tiredness and insecurity are examples of that. On top of that, a lot of the affected persons are scared that online attacks could be realized and end up in physical violence. As a consequence, they minimize their use of social media, delete their accounts or partly shut off their social life to protect themselves.

Fake news also has a high impact on society. A great example of that is the pizza-gate case. After voices got loud, that Donald Trump had bragged about sexually assaulting women, he pointed out that Bill Clinton had raped women and wanted to focus the attention on those accusations.¹⁰⁷ The leaked E-Mails of Hillary Clinton were said to prove that the rape shall have taken place repeatedly at the “Comet Ping Pong” Pizzeria in Washington D.C. According to the fake news, this Pizzeria was a meeting point of pedophiles, who could order “Pizza”, if they wanted to be served a girl, a “Hot Dog”, if they wanted to be served a boy and “Sauce” for an orgy.¹⁰⁸ There was no evidence that those meetings took place, but people were very interested in this scandalous story and therefore, the information was spread rapidly online all over America and even across the borders. The backfire of all the shocked citizens even went so far, that the owner of the Pizzeria, James Alefantis, received several death threats and the FBI had to take over this case.

As a consequence of the overwhelming fake news-flood on the pizza-gate case, something terrible happened on the 4th of December 2016. 28-year-old Edgar W. decided to travel from North Carolina to Washington D.C. with the intent of investigating the pedophile meetings by himself. He wanted to help the children that were being held hostage and free them from their suffering. When arriving at the Pizzeria, he brought a gun with him and was overwhelmed by his emotions. In a stroke of anger, he opened the fire and was shooting around with no specific target.¹⁰⁹

¹⁰⁶ <https://www.lmz-bw.de/medien-und-bildung/jugendmedienschutz/hatespeech/folgen-fuer-den-gesellschaftlichen-zusammenhalt/> (last accessed 07.11.2021)

¹⁰⁷ <https://www.faz.net/aktuell/feuilleton/debatten/wie-sich-in-amerika-die-herrschaft-der-luege-festigt-14565557.html> (last accessed 08.11.2021)

¹⁰⁸ *ibid.* (last accessed 09.11.2021)

¹⁰⁹ <https://www.faz.net/aktuell/gesellschaft/kriminalitaet/pizzagate-fall-mann-kriegt-4-jahre-haft-wegen-selbstjustiz-15073545.html> (last accessed 09.11.2021).

Fortunately, nobody was hurt. After police arrived, Edgar W. was arrested and then sentenced to 4 years in prison.

Taking a closer look at this case demonstrates, how dangerous fake news can be and what it can lead to. Here, they were once virtual theories with no truthful evidence and ended up becoming reality for some citizens.

What is special in this scenario is, that Edgar W. didn't want to hurt anybody, he just wanted to free the kids and protect them from more misery. Normally, this would be a heroic act but instead of becoming a hero, this man ended up becoming a criminal and almost a murderer. He has to spend 4 years in prison, might lose contact with various people in his life and might not be able to fully recover (both, emotionally and socially) from everything after he is released from prison.

What took place on the 4th of December 2016 is a perfect example for showing that not only people with bad ideas and criminal thoughts are receptive to fake news, but also people who are willing to help others and make a change for the better. Therefore, it is important to understand how and why some people are more receptive to fake news than others, which will be discussed in the next chapter.

2.5. Why do people fall for fake news?

Why do they not recognize fake news when they see them? Unfortunately, part of the problem with fake news is that people fall for it even when confronted with fact checks. Psychologists and other social scientists are working hard to understand the mechanisms behind the human mindset. There are psychological aspects that explain the mechanisms behind it; because our brain is a very powerful organ that sometimes tries to take efficient shortcuts. These shortcuts are also called heuristics - they are mental strategies and rules of thumb that help us make decisions and judgments with limited knowledge and time. Heuristics and cognitive biases can be dangerous because they can lead us to have unrealistic expectations and make poor decisions. Cognitive biases are gaps in reasoning, remembering, or evaluating something that can lead to false conclusions. They are universal and everyone has them. These aspects contribute to the success of fake news: We tend to avoid cognitive dissonance and prefer information that fits our worldview and social identity. The world and its social and economic interrelationships are complex and therefore simple explanations and/or solutions are sometimes unconsciously preferred. This applies to fake news just as much as to serious news.¹¹⁰

A cognitive bias is a subconscious error in thinking that leads you to misinterpret information from the world around you and affects the rationality and accuracy of decisions and judgments. Biases are unconscious and automatic processes designed to make decision-making quicker and more efficient. Cognitive biases can be caused by several different things, such as heuristics (mental shortcuts), social pressures, and emotions.¹¹¹

This chapter is focused on how cognitive biases work. It attempts to address the question of what goes on in people's minds that makes us more susceptible to falling for fake news and believing misinformation even after it has been corrected? The roles of mass media, as well as social media platforms in the spread of fake news, are also discussed.

¹¹⁰ <https://www.cits.ucsb.edu/fake-news/why-we-fall> (last accessed 24.11.2021).

¹¹¹ <https://www.simplypsychology.org/cognitive-bias.html#definition> (last accessed 09.11.2021).

2.5.1. Cognitive bias

Numerous cognitive biases influence our thinking in general, but also the way we think about others. They affect the decisions we make, as well as how we use information. The following figure shows the most common cognitive biases to be aware of. As mentioned earlier, everyone has these biases, and the first step to reducing their influence on them is to be aware of them.¹¹²

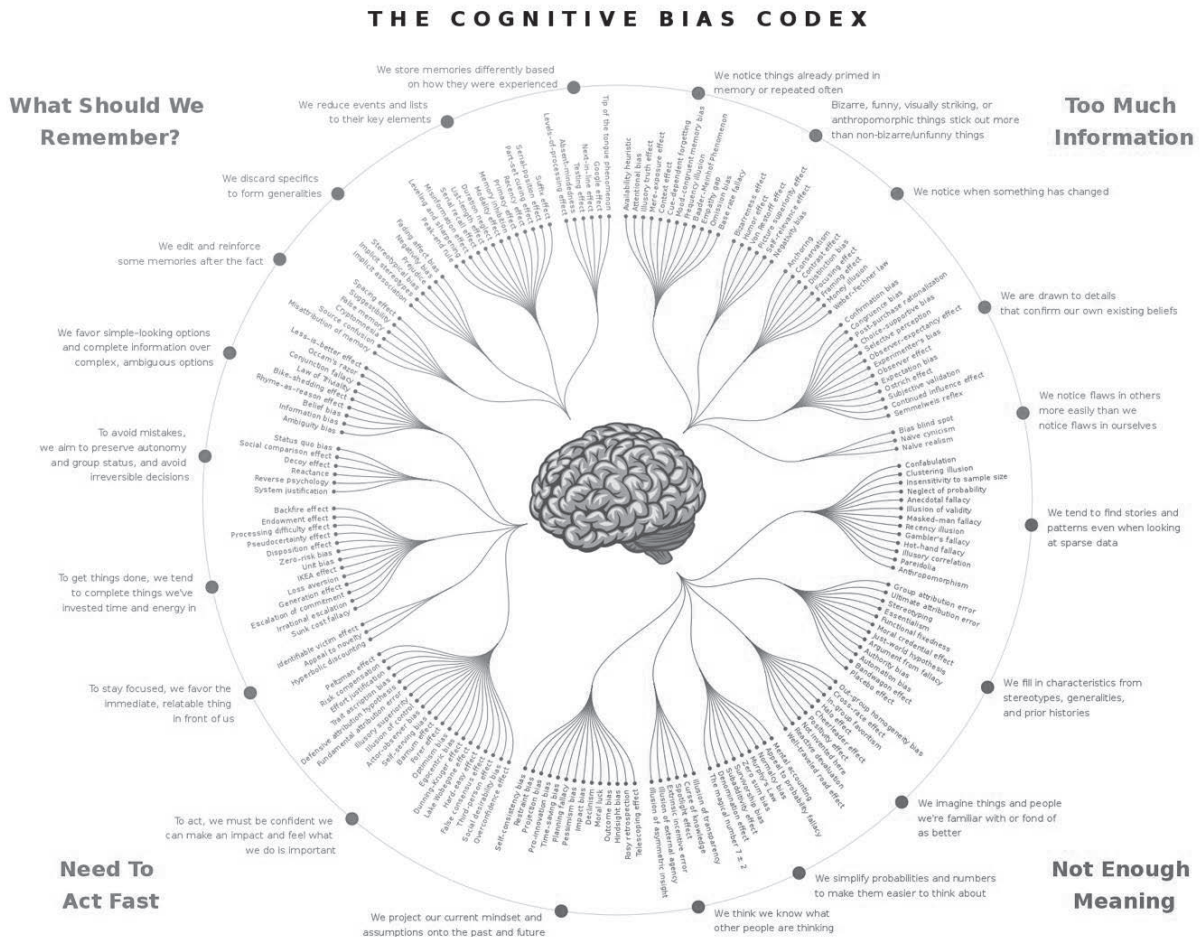


Figure 16: The Cognitive Bias Codex - 180+ biases.¹¹³

The following describes four types of cognitive biases that are particularly relevant in the context of fake news and its influence on society and politics.¹¹⁴

1. Reasoning: Reading and thinking is fundamental

An uncomfortable fact about social media is that people tend to form opinions about the news by paying attention only to the headline, without even reading the accompanying information and without evaluating it thoroughly afterward. A disturbing, if amusing, an example of this is a social experiment conducted by National Public Radio (NPR): "Why Doesn't America Read Anymore?"

¹¹² <https://www.cits.ucsb.edu/fake-news/why-we-fall> (last accessed 24.11.2021).

¹¹³ <https://commons.wikimedia.org/w/index.php?curid=69756809> (last accessed 20.01.2022).

¹¹⁴ <https://www.cits.ucsb.edu/fake-news/why-we-fall> (last accessed 24.11.2021).

was the post that NPR jokingly placed on its Facebook page. When users followed the linked-up post, they were directed to the article on the NPR website, which explained that the post was a hoax. Nevertheless, many viewers did not read the clarifying article and responded immediately with comments on the Facebook page.¹¹⁵

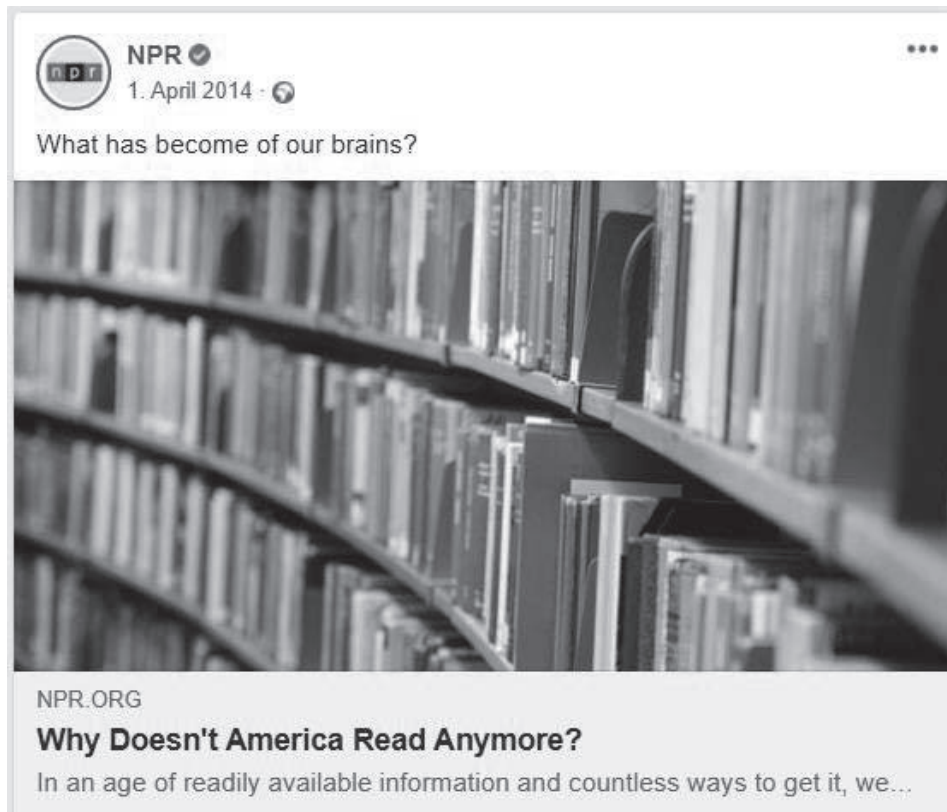


Figure 17: NPR on Facebook “What has become of our brains?”.¹¹⁶

Other news providers also report similar experiences. Researchers have studied this problem and examined 2.8 million news articles on Twitter that were shared or commented on by its users. According to the results, more than half of these users never clicked on the shared link and read the entire article. So, people often share or like headlines online without having engaged with the actual content [2-21, 2-22]. This behavior could be very harmful in terms of the spread and influence of fake news. Likewise, it leads to the rise of clickbait as the eye-catching headlines attract attention [2-23]. Thus, people may mistakenly assume that the misleading headlines are true when they do not read the content and do not further explore whether there is any doubt about the story or any other opinion. Sharing headlines without reading the content can also make it appear that they are gaining popularity, i.e., are trending [2-24]. This increases the likelihood that other people will continue to share these headlines without having read them, leading to a kind of socio-cognitive epidemic.

Researchers from the Massachusetts Institute of Technology (MIT), USA, and the University of Regina in Saskatchewan, Canada, studied the analytical reasoning ability of more than three thousand American participants. Their results showed a correlation between a higher score on the intelligence test and a better ability to distinguish fake news headlines from real news headlines. This was the

¹¹⁵ https://www.npr.org/2014/04/01/297690717/why-doesnt-america-read-anymore?utm_medium=facebook&utm_source=npr&utm_campaign=nprnews&utm_content=04012014 (last accessed 24.11.2021).

¹¹⁶ https://www.facebook.com/NPR/posts/10202059501509428?stream_ref=10 (last accessed 24.11.2021).

case even when the fake news matched the political preferences of the participants. According to the results of the authors, people were more likely to fall for fake news because of laziness of thinking than because of a conscious or unconscious desire to confirm their political preference [2-25]. “People [who] believed false headlines tended to be the people [who] didn’t think carefully, regardless of whether those headlines aligned with their ideology,” Rand said [2-26]. Confirming this conclusion, a replication study by Rand with colleagues showed that careful thinking had an effect not only on the detection of patently false headlines but also on nonpartisan headlines [2-27]. These results were confirmed in further research. Study participants were shown a series of news stories as they would appear in social media: as screenshots with the headline, source and first sentences of a news story. Some of these stories were false; others were true. Participants were first asked to quickly and visually judge whether the news stories were real or fake. Later, they were asked to judge these stories again, but this time they were asked to take more time and think carefully about the truthfulness of the news. The comparison between the fast and slow procedures showed that participants were better able to discern the truth content when they took more time to think, regardless of the political consistency of the news they evaluate [2-28].

In summary, the (in)ability to distinguish between truth and untruth is related to the ability to reflect on the content read (or heard). Accordingly, more reflective, analytically thinking individuals are less inclined to fall for fake news. Belief in fake news is likewise related to delusionality, dogmatism, religious fundamentalism [2-29], bullshit receptivity, and overclaiming [2-30] – all of which are likewise factors related to analytical thinking.

2. Bandwagon effect: Thousands of “likes” and “shares” cannot be wrong

The apparent popularity of a news story fuels another bias and promotes the acceptance of fake news. The so-called “bandwagon effect” is one of the most researched cognitive biases and explains the influence of popularity on our thinking, e.g. [2-31]. According to this, people tend to adopt a certain behavior, in this case sharing news, simply because others are already doing it. So the more people adopt a particular trend, the more likely it is that other people will also “get on the bandwagon” [2-32]. To this end, the brain uses mental shortcuts (heuristics) designed to facilitate rapid decision-making. It usually takes time to think through an idea or behavior before implementing it. Many people skip the process of individual evaluation by relying on the judgment of others around them whose opinions are perceived as trustworthy - and so the third-party opinion gains popularity.

The bandwagon effect leads us to be more influenced by what has been shared and liked, rather than focusing on what is actually in the content. Related to fake news, this effect in a sense frees us from the responsibility of having to verify the information shared. We simply assume that someone must have verified the information if it has already been shared many times [2-22]. To confirm this assumption, the researchers manually reviewed a sample of 50 articles shared on Twitter, drawn evenly at random from a corpus of articles. The result showed that only a minority of the shared articles contained verified claims:

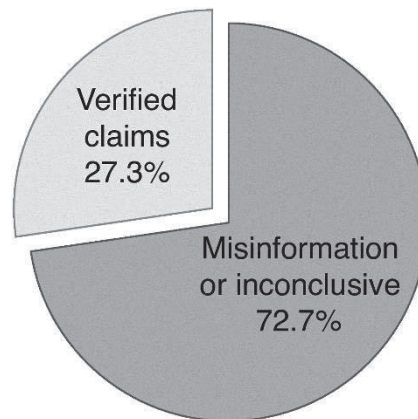


Figure 18: Verification based on a sample of 50 articles.

Source: [2-22]

The bandwagon effect is a similar phenomenon to "herd mentality" or "groupthink". The cause may be the pressure to conform when it appears that a great many people agree with a message or an opinion [2-33]. Further, the human desire to be right and appear to be an expert may be a reason why many people claim and share information that they have not even read. We all want to be on the winning side and tend to look to other people in our social group to find out what is right and acceptable and adjust our (political) thinking accordingly [2-34]. Likewise, the need to be included can play a role in the acceptance and spread of fake news. Generally, people do not want to be seen as outsiders but want to be liked and go along with what people around them are doing to secure inclusion and social acceptance [2-35]. People want to feel connected to those around them, whether it's members of the same political party, political activists highlighting climate change, members of religious communities, etc.

The bandwagon effect can have several negative effects. For example:

- Researchers have found that polls can influence individuals when a particular candidate is ahead in them. People may tend to then change their vote to be on the winning side. [2-36, 2-37].
- Individuals have been influenced by the anti-vaccination movement and, as a result, were less likely to have their children routinely vaccinated. This movement has been linked to the recent measles outbreak [2-38]. This example can currently be applied to the COVID-19 vaccination campaign as well.

Social media platforms have an important part in the emergence of the bandwagon effect. In relation to social media, the problem of the influence of popularity on the acceptance of fake news is amplified by so-called "social bots". Bots are programs that perform tasks automatically. They can be harmless or even useful or amusing. For example, bots that automatically answer frequently asked questions (FAQ) in a chat are useful. However, bots on social media can also be used with malicious intent to manipulate users by spreading false information or making it appear that certain people, ideas, or products are more popular than they actually are. They can also troll or attack victims and commit many other types of abuse. In the broadest sense, a social bot is a social media account that is controlled, at least in part, by software. The content of posts and the timing of actions usually originate from humans or are closely monitored by humans. Most automated bots simply act (reply, post,

follow, etc.) based on triggers or scripted patterns, such as retweeting all messages from specific accounts or posting from predefined lists.¹¹⁷

3. Partisanship: The most common political mental shortcut

Our personality and worldview can also lead us into mental traps. Especially in the case of fake news, our political attitude has a significant part in it, and it is not easy to avoid these traps. Regardless of its orientation, political preferences can affect whether we believe or reject the news; the actual truthfulness does not matter. Research findings supported the thesis that most politically oriented fake news sites were consumed by conservatives. For example, surveys and internet traffic data from the 2016 U.S. presidential campaign showed that Donald Trump's supporters in particular were most likely to visit untrustworthy websites that were frequently shared via Facebook [2-39, pp. 12-18].

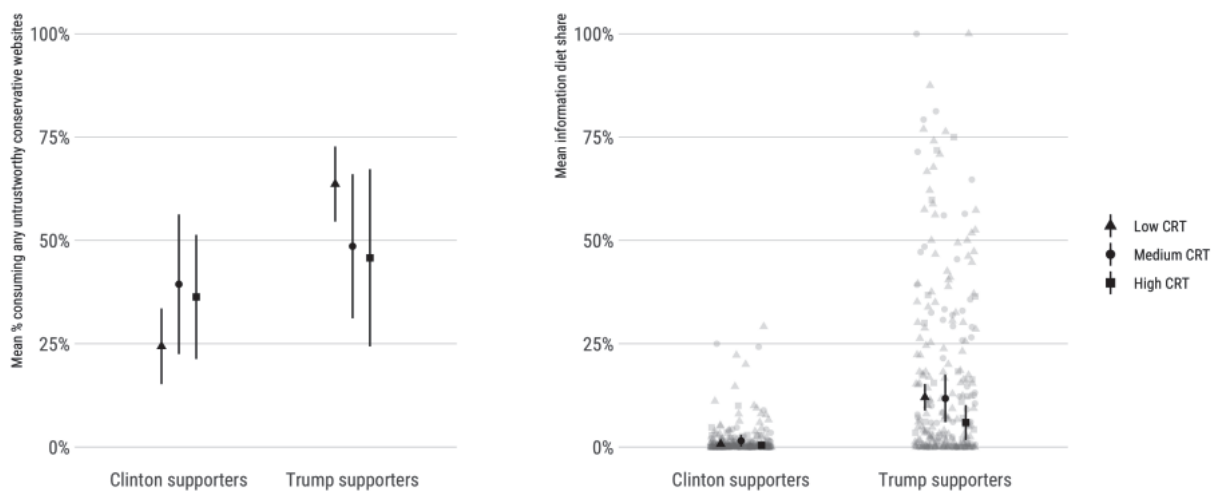


Figure 19: Consumption of untrustworthy conservative websites by CRT score and candidate preference.

Source: [2-39, p. 13]

The collected data also revealed that supporters of presidential candidate Hilary Clinton on average visited fact-checking websites more often and fake news websites less often. The opposite was observed for Trump supporters [2-39, pp. 12-18].

There is a widespread view that the inability to distinguish between true and false news is due to political motivations. When people are confronted with politically valued content, they display an “identity-protective perception.” Thus, this results in people believing such content that is consistent with their partisan political identity. In contrast, they in turn give little credence to such content that is inconsistent with their political orientation [2-40, 2-41]. A related theory suggests that people place loyalty to their political identity above truth – and therefore fail to distinguish truth from falsehood and instead simply believe ideologically consistent information [2-42]. These explanations assume that a strong causal influence of political motivation on belief is the dominant factor explaining why people fall for fake news [2-43].

Individuals whose business concept is to make money by spreading fake news on the Internet stated that they tried to target both conservatives and liberals with false news, regardless of their political

¹¹⁷ <https://botometer.osome.iu.edu/faq> (last accessed 29.11.2021).

preferences. However, people with liberal political views were less likely to click on these stories, which led such websites to stop disseminating pro-liberal fake news:

“Well, this isn’t just a Trump-supporter problem. This is a right-wing issue. Sarah Palin’s famous blasting of the lamestream media is a kind of record and testament to the rise of these kinds of people. The post-fact era is what I would refer to it as. This isn’t something that started with Trump. This is something that’s been in the works for a while. His whole campaign was this thing of discrediting mainstream media sources, which is one of those dog whistles to his supporters. When we were coming up with headlines it’s always kind of about the red meat. Trump got into the red meat. He knew who his base was. He knew how to feed them a constant diet of this red meat. We’ve tried to do similar things to liberals. It just has never worked, it never takes off. You’ll get debunked within the first two comments and then the whole thing just kind of fizzles out.” [2-44]

Perhaps the clearest evidence that Trump knew exactly how to “feed” his base is the violent storming of Capitol Hill in Washington, D.C., by an estimated 2,000 to 2,500 of his supporters on January 6th, 2021 – just minutes after Trump explicitly asked them to march there. These riots have so far been the culmination of Trump’s unsuccessful claims of voter fraud in the November 2020 presidential election and repeated stoking of division in the United States. In a constant barrage of misinformation, Trump repeatedly claimed months before the election that he would only lose if the election was rigged [2-45].

Generally speaking, the political landscape seems to be increasingly populated by actors who spread demonstrably false claims. This problem is exacerbated by foreign efforts, such as Russian propaganda tactics, including on social media, that attempt to influence elections or influence misinformation regarding the COVID-19 pandemic [2-46, pp. 5-6].

4. Belief echoes: The tendency for misinformation to persist

What is worrisome about cognitive bias is that it can persist for a long time, preventing us from eliminating misinformation that has already been internalized. Unfortunately, it is not the case that belief in fake news can be quickly eliminated simply when that news has been disproved or corrected.

Scientific research has confirmed that our memory is not good at remembering what is real and what is fake once we receive the information. Professor Emily Thorson at Boston College has found that even when misinformation has been corrected, “belief echoes” remain. Her research suggests that these echoes can be generated by an automatic or deliberative process: belief echoes occur even when the misinformation is immediately corrected according to the “gold standard” of journalistic fact-checking [2-47].

The “United Kingdom European Union membership referendum” on June 23rd, 2016 – the so-called Brexit – can serve as a vivid example of how the tendency for misinformation to persist can influence political reality. One of the central messages of Brexit supporters in the run-up to the referendum was that Britain was paying 350 million GBP (just under 400 million Euro) each week to fund the budget of the European Union¹¹⁸. This claim was circulated by two key figures in the Brexit campaign – Boris Johnson and Nigel Farage – who stated that this money would be better spent on the British healthcare system (NHS National Health Service) after all. Before the referendum, various institutions examined

¹¹⁸ http://www.voteleavetakecontrol.org/briefing_cost.html (last accessed 13.12.2021).

the claims and consistently reported that the amount stated was incorrect.¹¹⁹ The day after the referendum in which Britons voted for Brexit, Nigel Farage even publicly admitted the “mistake”.¹²⁰

Even if some forms of fake news warning pointed people not to believe it, the absence of such warnings can have a greater impact than their presence. Scientists have addressed the question of what can be done to counter political misinformation: they studied how fake news warnings affect people’s beliefs. As a result, people were less likely to believe such stories when they came with a warning; but conversely, when a warning was not present, subjects were more likely to believe the stories, whether they were fake or not. Thus, when people are confronted with a warning about misinformation, they are more likely to feel they do not need to be on alert and to rely on the warning. However, in the absence of a warning about misinformation, people are more inclined to believe the information to be credible, even though it is not. Taken together, these results challenge theories of motivated reasoning while highlighting a potential challenge to the politics of using warnings to fight misinformation – a worrying challenge because it is much easier to produce misinformation than to debunk it [2-48].

Twitter-related research has shown that users who posted misleading content often tagged it with a fact-check article [2-22]. The problem, however, was that even when people clicked on the fact check (which doesn’t always happen – such as “bandwagon effect”), the facts cited were themselves fake, which is often exploited by clickbait creators to generate clicks and make money that way [2-23].

Doubts should also be raised about the use of warnings, and they should be chosen carefully. If they are formulated too drastically, they could cause a reaction. If they are formulated too weakly, they could be overlooked. Moreover, they could be forgotten over longer periods. Nevertheless, they cannot be dispensed with completely. In particular, warnings against further dissemination, i.e., sharing of false news, appear to be important, since personal recommendation from user to the user represents one of the central mechanisms in the spread of false news on the Internet [2-49].

2.5.2. The role of mass media

As Heinz Bonfadelli of the German Federal Agency for Civic Education (bpb – Bundeszentrale für politische Bildung) sums it up well, mass media such as the press, radio and television, as well as the Internet and social web, make an indispensable contribution to the functioning of democracy. Politicians in general and media professionals in particular, but also the public, assume this. Mass media are supposed to contribute to both the stability and the change of society [2-50].

According to sociologist Niklas Luhmann [2-51], media enable society to observe itself:

- Media as “windows to the world” select and provide relevant topics for the public.
- Media provide citizens with arguments for and against controversial issues.
- Media research the background knowledge necessary for decision-making, prepare it in a comprehensible way and make it widely available.

As a result of this media service, arguments on current issues are exchanged, discussed and critically scrutinized in public. By using the media, the population participates in current social issues and problems. This increases the level of knowledge of all. In addition, it is hoped that minorities such as migrants will also be integrated into society through the media. Through media coverage, social prejudices and perhaps even discrimination against minorities could weaken [2-50].

¹¹⁹ e.g., <http://infacts.org/quitting-eu-mean-less-money-nhs-not/>, also <https://www.bbc.com/news/uk-politics-eu-referendum-36110822> (last accessed 13.12.2021).

¹²⁰ <https://www.itv.com/goodmorningbritain/articles/nigel-farage-labels-350m-nhs-promise-a-mistake> (last accessed 13.12.2021).

Information as a “window on the world” → transparent public arena		
Politics	Economy	Culture – Social
Creating publicity	Consumer information	Orientation and life support
Articulation of opinions	Circulation of goods	Socialization: values & norms
Control and critique	Securing employment	Integration in society
Early warning function	Value creation, e.g., in the media industry	Education and cultural development
Participation & activation		Entertainment and relaxation

Figure 20: Overview: Functions of the media for the sectors of society.

Source: [2-50]

These expectations of the public mass media are ideal concepts, which are demanded as desirable achievements. In reality, however, they are always only partially realized, which is repeatedly expressed in media criticism and media scolding. Instead of transparent diversity of opinion, the opinion of the government or powerful groups can dominate the media unquestioned as a uniform majority opinion, especially in authoritarian societies with limited media freedom (e.g., Russia or China). But the question also arises for media in democracies as to whether and to what extent they are concretely committed to more or less equality in society. Instead of contributing to integration and solidarity about migrants or other minorities, media can contribute to stereotyping and reinforce discrimination through blanket negative reporting. Finally, there is always the danger of unjustifiably discrediting individuals or social groups through one-sided moralizing portrayals. Content analyses of (German, editor’s note) media coverage show that migrants and especially Muslims tend to feature little in the media, and when they do, they are portrayed stereotypically and negatively [2-50, 2-52].

Events of the recent past in the media sector give cause for concern that the quality of media reporting is in danger. Warning voices even speak of a media crisis (cf. [2-53]). In the print sector as well as in broadcasting, a growing media concentration has been underway among media groups for some time: large media groups are becoming increasingly dominant. In parallel, advertising spending is shifting from the press to the Internet, and newspaper use is declining. At the level of media organizations, this has led not least to the layoff of media professionals, the downsizing of newsrooms, and the creation of lower-cost newsrooms. In the newsroom, content is produced jointly for the print edition and the online offering. Journalists thus no longer write an article just for the newspaper, but also create online versions or radio or TV reports at the same time. This has led not least to an increase in the time pressure of journalistic work. But the media crisis is not just a funding crisis; journalism is also affected in terms of content. Commercialization has not only led to a decrease in media diversity but the economic pressure is also expressed in an increased external influence of public relations on reporting, for example as courtesy journalism. The blurring of the boundaries between editorial and advertising sections (keyword: native advertising) endangers journalistic independence. As a result of economization, there is also an increased focus on the audience and its wishes. Information and entertainment, as well as the public and private spheres of politicians, for example, are mixed in reporting to make it more interesting. Criticism of the media focuses on the tabloid press on the one hand and private broadcasting on the other under the headings of “personalization” [2-54] and “infotainment” [2-55]. Both are accused of populism and a lack of independence, as well as a generally low level of quality [2-50].

Infotainment / Personalization

The mixing of informative and entertaining formats of television is called infotainment. The first part of the word comes from “information”, the second part is derived from the Anglo-American term “entertainment”. As a rule, this describes the tendency, for example, to include more and more “soft” topics such as news about celebrities in news programs. Infotainment also refers to the increasing emotionalization and personalization of news, the latter meaning the focus on a specific person (presenter, “anchorman”).¹²¹

Current crises, such as the COVID-19 pandemic or climate change, have brought about another development that, while not unknown, is putting media ethics to the test. This is the journalistic professional standard of reporting news in as balanced and neutral a manner as possible to also allow for nuanced voice. Journalists adhere to this norm to demonstrate their professionalism and avoid any criticism of their one-sided reports. At the same time, balance can also substitute for plausibility checks, as when reporters do not have enough time for research, or their own skills are not sufficient to assess the validity of certain contradictory statements. When controversial statements are combined with reporters' lack of expertise, the standards of balance become particularly apparent. However, when the voices of dissenting outsiders are copied out of context, it provides them with legitimacy and media prestige that may also enable them to gain political power. Research also shows that ideological bias can also play an important role: For example, right-wing and conservative columnists predominantly allow climate change deniers to have their say in their articles [2-56].

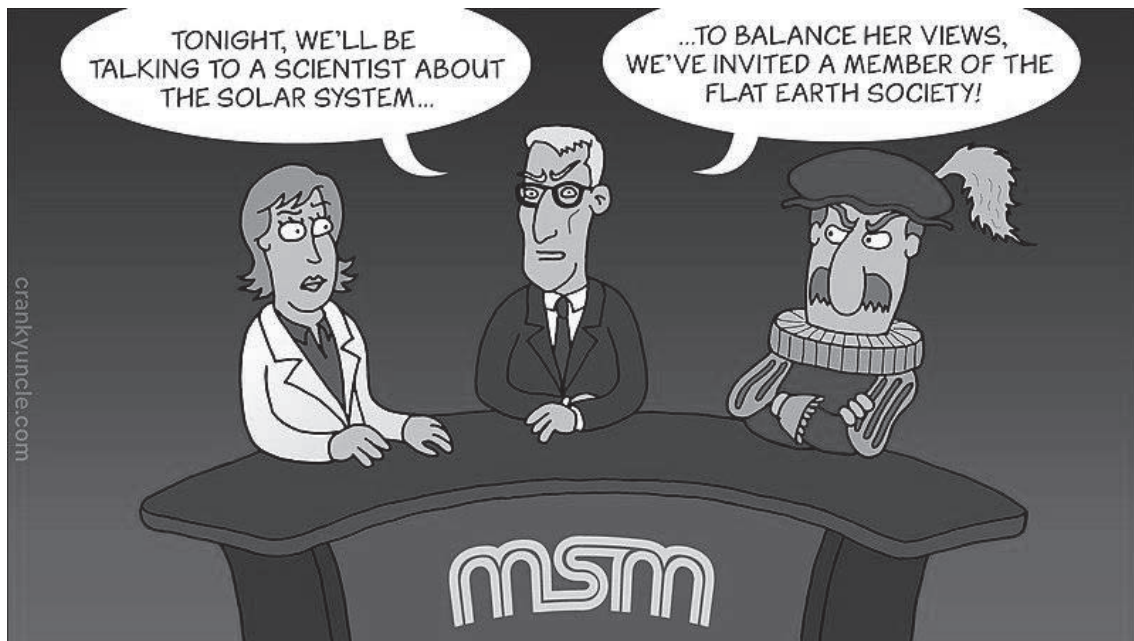


Figure 21: Example of false balance media coverage.¹²²

Especially in the context of scientific reporting on the causes and combating of current crises, the term “false balance” has become more popular due to the journalistic practice described above. While balanced and independent reporting is generally considered good journalism, and this balance does have its purpose. The public should always be able to trust that all important aspects will be published in the media, and everyone should be able to find out about all sides of an issue. At the same time,

¹²¹ [2-50], own translation.

¹²² https://commons.wikimedia.org/wiki/File:False_balance.jpg (last accessed 21.01.2022).

this should not mean that all sides of an issue deserve equal weight either - e.g., one guest “pro” and one “contra”. Science is based on evidence for different hypotheses, which are carefully tested and then built upon those that provide the most evidence. However, when all different opinions and viewpoints are presented side by side and the same proportion in the press or in politics, the impression of equal weight scientific legitimacy is falsely created. In this way, one of the main goals of science is defeated: weighing the evidence.¹²³

False balance

is a misleading argument that presents two or more positions as equally valid when the evidence strongly supports one over the others. This may be unintentional faulty reasoning. Alternatively, a false balance may be used to influence and mislead.¹²⁴

Moreover, polarizing topics are particularly favored by the media. For example, in the area of vaccinations against the coronavirus (COVID-19), this seemingly balanced presentation contradicts both the expert consensus expressed in the vaccination recommendations of the European Medicines Agency (EMA)¹²⁵ and the national control authorities, such as the German STIKO,¹²⁶ and the distribution of opinion in the population [2-57], thus falsely suggesting an equal distribution of opinions. Such false balance leads to uncertainty [2-58] and raises doubts in the population and favors the emergence of fake news, especially in social media. Making the media and their reporting aware of this and calling on them to weigh the evidence rather than the opinions is also an important contribution to strengthening public confidence in vaccination, researchers urge [2-59, pp. 404-405].

The project “Understanding Science” by the University of California Museum of Paleontology, Berkeley, has summarized six steps that can be applied to evaluate scientific messages:¹²⁷

1. Where does the information come from?
 - What is the source of this message?
 - Does that source have an agenda or goal?
2. Are the views of the scientific community accurately portrayed?
 - Who is the expert?
 - Beware of false balance.
3. Is the scientific community's confidence in the ideas accurately portrayed?
 - Science is always ready to revise ideas when new evidence justifies them.
 - Tentativeness does not mean that scientific ideas are untrustworthy ... and this is where some media reports on science can mislead, mistaking normal scientific proceedings for untrustworthiness.
4. Is a controversy misrepresented or blown out of proportion?
 - Fundamental scientific controversy: scientists disagreeing about a central hypothesis or theory.

¹²³ https://undsci.berkeley.edu/article/sciencetoolkit_04 (last accessed 04.01.2022).

¹²⁴ <https://simplicable.com/new/false-balance> (last accessed 04.01.2022).

¹²⁵ <https://www.ema.europa.eu/en/human-regulatory/overview/public-health-threats/coronavirus-disease-covid-19/treatments-vaccines/vaccines-covid-19/covid-19-vaccines-authorized> (last accessed 04.01.2022).

¹²⁶ <https://www.rki.de/DE/Content/Infekt/Impfen/ImpfungenAZ/COVID-19/Impfempfehlung-Zusfassung.html> (last accessed 04.01.2022).

¹²⁷ https://undsci.berkeley.edu/article/0_0_0/sciencetoolkit_02 (last accessed 04.01.2022).

- Secondary scientific controversy: scientists disagree about a less central aspect of a scientific idea.
- Conflict over the ethicality of methods: disagreement within the scientific community or society at large over the appropriateness of a method used for scientific research.
- Conflict over applications: conflict over the application of scientific knowledge.
- The conflict between scientific ideas and non-scientific viewpoints.

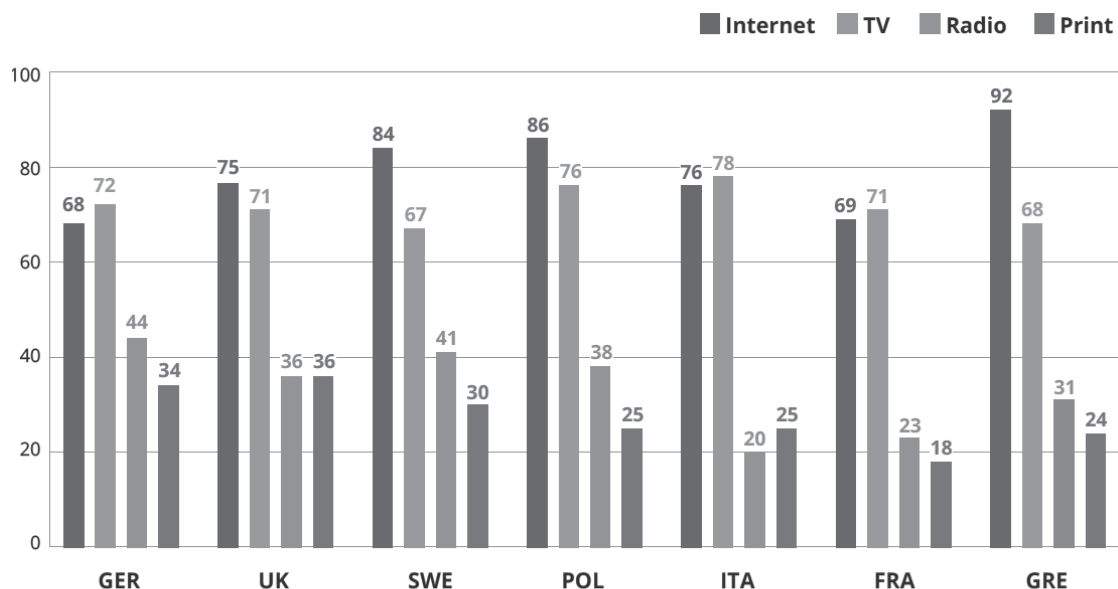
5. Where can I get more information?

- Find sources with scientific expertise.
- Avoid ulterior motives.
- Keep it current.
- Check for citations.

6. How strong is the evidence?

- Does the evidence suggest correlation or causation? In other words, do the data suggest that two factors (e.g., high blood pressure and heart attack rates) are correlated with one another or that changes in one cause changes in the other?
- Is the evidence based on a large sample of observations (e.g., 10,000 patients with high blood pressure) or just a few isolated incidents?
- Does the evidence back up all the claims made in the article (e.g., about the cause of heart attacks, a new blood pressure drug, and preventative strategies) or just a few of them?
- Are the claims in the article supported by multiple lines of evidence (e.g., from clinical trials, epidemiological studies, and animal studies)?
- Does the scientific community find the evidence convincing?

2.5.3. The role of social media platforms



Source: Reuters Institute Digital News Survey 2019

Bases: GER: n = 2022; UK: n = 2023; SWE: n = 2007; POL: n = 2009; ITA: n = 2006; FRA: n = 2005; GRE: n = 2018

Figure 22: Regularly used news sources 2019 (Percentage).

Source: [2-60, p. 18]

Unlike in the past, when information was only obtained from mass media such as newspapers, television, and radio, today we are constantly flooded with information. Social media platforms (Twitter, Facebook, Instagram, YouTube, etc.) and messenger services (Telegram, WhatsApp, etc.) have grown immensely in importance and now play a key role in the spread of information. But not only news is shared on social media - regardless of its truth content - but also opinions. The problem with this is that opinions often do not necessarily match the facts. No matter how convincing facts may be, they do not always influence the opinion once it has been formed. Scientists have known about the phenomenon of unbending personal opinion for decades (e.g. [2-61, 2-62]).

This phenomenon of unbending personal opinion is known as *confirmation bias*. Confirmation bias is “the tendency to gather evidence that confirms preexisting expectations, typically by emphasizing or pursuing supporting evidence while dismissing or failing to seek contradictory evidence”¹²⁸. In other words, it describes the fact that we generally regard information that supports our worldview as credible and relevant, while we ignore or dismiss as nonsense information that does not fit into our worldview.

The developers of social media platforms and online search engines are well informed about how human cognitive biases work. With this knowledge, personalized technologies - called algorithms - are created to determine what people should see online. These personalization technologies are built to pick only the most interesting and relevant content for each user, and can thus reinforce users’ cognitive and social biases and make them more susceptible to misinformation and fake news. The detailed advertising tools built into many social media platforms, for example, allow disinformation activists to exploit confirmation bias by tailoring messages to people who are already inclined to believe them. If a user frequently clicks on Facebook links from a particular news source, Facebook tends to show that person more content from that source. This creates what is known as the “filter bubble” or “echo chamber” effect [2-63]. This effect can isolate people from other perspectives and reinforce confirmation bias [2-64].

Another important component of social media is information trending on the platform based on what gets the most clicks. Authors and researchers Giovanni Luca Ciampaglia and Filippo Menczer refer to this as popularity bias. In their research, they found that an algorithm that aims to promote popular content can have a negative impact on the overall quality of information on the platform. This similarly impacts existing cognitive biases, as what appears to be popular is promoted regardless of its quality. Such algorithmic biases can be manipulated by the effect of the social bots [2-65]. These computer programs interact with humans via social media accounts and most, such as Big Ben¹²⁹ from Twitter, are harmless. However, there are also social bots in use that hide their true nature and serve malicious purposes, such as spreading disinformation or falsely creating the appearance of a grassroots movement¹³⁰, also called “astroturfing”¹³¹ [2-64].

To study how the structure of online social networks makes users vulnerable to disinformation, the Hoaxy system was developed.¹³² Hoaxy tracks the spread of content from sources with low credibility visualizes it and shows how it competes with fact-checking content. Analysis of data collected by Hoaxy during the 2016 U.S. presidential election found that Twitter accounts that spread misinformation were almost completely cut off from fact-checkers corrections [2-22]. Accounts that

¹²⁸ <https://dictionary.apa.org/confirmation-bias> (last accessed 27.10.2021).

¹²⁹ https://twitter.com/big_ben_clock (last accessed 04.01.2022).

¹³⁰ e.g., <https://www.businessinsider.com/astroturfing-grassroots-movements-2011-9> (last accessed 04.01.2022).

¹³¹ <https://lobbypedia.de/wiki/Astroturfing> or <https://www.merriam-webster.com/dictionary/astroturfing> (last accessed 04.01.2022).

¹³² <https://hoaxy.osome.iu.edu> (last accessed 04.01.2022).

spread misinformation were examined more closely. The authors found a very dense core group of accounts that retweeted each other almost exclusively - including several bots [2-64].

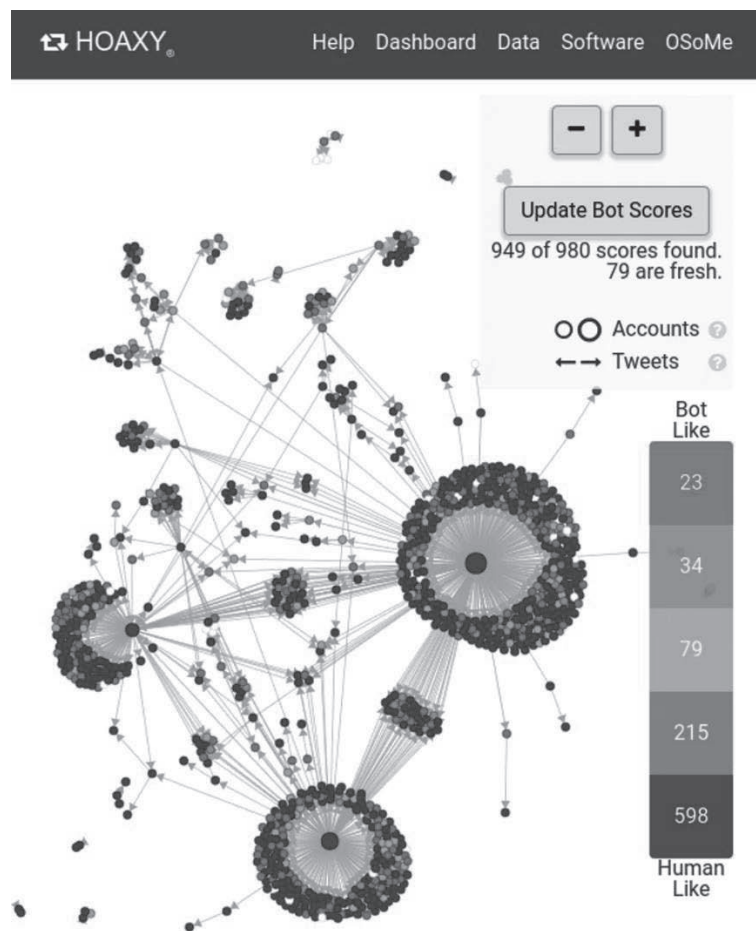


Figure 23: A screenshot of a Hoaxy search shows how common bots – in red and dark pink – are spreading a false story on Twitter.¹³³

As a collaborative project of the Observatory on Social Media (OSoMe, pronounced “awesome”)¹³⁴ at Indiana University, USA, the tool “Botometer”¹³⁵ was also developed. This tool helps to detect social bots on Twitter and investigate the manipulation strategies behind them. Botometer uses machine learning (a branch of artificial intelligence – editor’s note) to detect bot accounts by examining thousands of different characteristics of Twitter accounts, such as the timing of their posts, the frequency of their tweets, and the accounts they follow and retweet. It turned out that up to 15 percent of active Twitter accounts showed signs of bots [2-66]. In combination with Hoaxy, many bots have been discovered that exploit their victims’ cognitive, confirmation, and popularity preferences as well as Twitter’s algorithmic preferences. The social bots enable filter bubbles to be built around vulnerable users and submit fake news to them. The bots first attract the attention of users who, for example, support a particular politician by tweeting that candidate’s hashtags or mentioning/retweeting the person. In addition, the bots can reinforce false claims and slander

¹³³ <https://theconversation.com/misinformation-and-biases-infect-social-media-both-intentionally-and-accidentally-97148> (last accessed 04.01.2021).

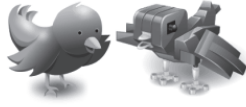
¹³⁴ <https://osome.iu.edu> (last accessed 04.01.2022).

¹³⁵ <https://botometer.osome.iu.edu> (last accessed 04.01.2022).

opponents by retweeting articles from less credible sources that match certain keywords. In this way, the algorithm also highlights false stories that are frequently shared with other users [2-64].

Botometer[®]

An OSoMe project (bot•o•meter)




Botometer (formerly BotOrNot) checks the activity of a Twitter account and gives it a score. Higher scores mean more bot-like activity.


Use of this service requires Twitter authentication and permissions. (Why?)

If something's not working or you have questions, please contact us only after reading the FAQ.


Botometer is a joint project of the Observatory on Social Media (OSoMe) and the Network Science Institute (IUNI) at Indiana University.

⚙️ @Skyjiv ✕

➤ @eat_sleep_talk is not active, score might be inaccurate 4.6 / 5  ✕

➤ @moreklm 0.8 / 5  ✕

➤ S @5471cul4 joined Twitter recently, score might be inaccurate ⚠️ No Timeline ✕

➤ @THE_REAL_JG 1 / 5  ✕

➤ @jschleeff ⚠️ unauthorized ✕

Figure 24: A screenshot of the Botometer website, which checks the followers of German Foreign Minister Annalena Baerbock for possible bot accounts.¹³⁶

As already discussed above, misinformation on the Internet has a direct impact on our society and its political, economic, and ideological orientation. The essential role of social media in the spread of fake news is currently receiving a lot of attention, and the political developments of recent years illustrate the pressing problem: Facebook and Cambridge Analytica, which significantly influenced the election of Donald Trump as president of the USA and likewise the Brexit referendum in 2016 [2-67] or the widespread skepticism about human influence on climate change,¹³⁷ for example.

In December 2021, it was announced that Rohingya who had fled the genocide in Myanmar was suing Facebook's parent company, Meta, for 150 billion US-Dollars in damages. They allege that Facebook's algorithm, which promoted extremism and violence, was knowingly and willfully used,

¹³⁶ Screenshot Botometer (last accessed 04.01.2022).

¹³⁷ <https://www.encyclopedia.com/environment/energy-government-and-defense-magazines/climate-change-skeptics> (last accessed 09.11.2021).

and that the corporation was complicit in the 2017 genocide. They claim that racist hatred of the Rohingya was deliberately fomented on the platform, misinformation was spread, and calls were made for violence against the predominantly Muslim ethnic group. “For example, the military called us ‘animals’, ‘monkeys’, ‘donkeys’, ‘rapists’ - this spread further and further on Facebook”, recalls Rohingya Ambia Perveen. An independent expert report confirms Facebook's complicity.¹³⁸

Already before the Rohingya lawsuit, Facebook’s involvement in spreading propaganda and fake news was highlighted by the revelations of former Facebook employee and whistleblower Frances Haugen. On the 4th of October 2021 hearing before the U.S. Senate, Haugen called for strict regulation of the global enterprise: It needs pressure, monitoring and oversight because so far Facebook has been a black box. Facebook's products harm children, stoke division in society, and weaken democracy. The company's leadership, she said, knows how to make Facebook and Instagram safer, but does not want to make the necessary adjustments because it puts its immense profits above the common good and the truth. “This has to change”, Haugen told the U.S. Senate, and for that reason, she decided to let the public know about it.¹³⁹

What does it mean that Facebook knows how to make its service safer? The answer to this question is hidden behind the algorithm used by Facebook. Facebook and every other social media service work with their algorithms. There is no clear definition of algorithms in social and cultural studies research. They are often understood as hidden and powerful mechanisms that have a great influence on our digital lives. The problem with algorithms stems from their lack of transparency, which is difficult for outsiders to comprehend - their detailed mode of operation is a closely guarded secret of social media providers [2-68, pp. 181-182]. Tarleton Gillespie describes the algorithm as a recipe assembled in programmable steps. Applied to social media providers, this, therefore, means that developers first define a model in which the problems to be solved and the goals to be achieved are formulated. Once such a model has been formulated, the recipe for it, i.e. the algorithm, is developed [2-69, p. 19].

From the perspective of social media users, this lack of transparency is particularly problematic because they do not know what content is being hidden from them by the algorithms. The systematic hiding of more differentiated content and the display of one-sided content can lead to a distorted view of the world, as described by the metaphors of the filter bubble and the echo chamber. This applies to political content as well as to ideologies and extreme viewpoints of all kinds [2-60, pp. 12-13].

¹³⁸ Ambia Perveen in conversation with Michael Borgers (text by Mike Herbstreuth): <https://www.deutschlandfunk.de/rohingya-klage-facebook-100.html> (last accessed 05.01.2022).

¹³⁹ statement of Frances Haugen: <https://www.washingtonpost.com/context/facebook-whistleblower-frances-haugen-s-senate-testimony/8d324185-d725-4d99-9160-9ce9e13f58a3/> (last accessed 09.11.2021).

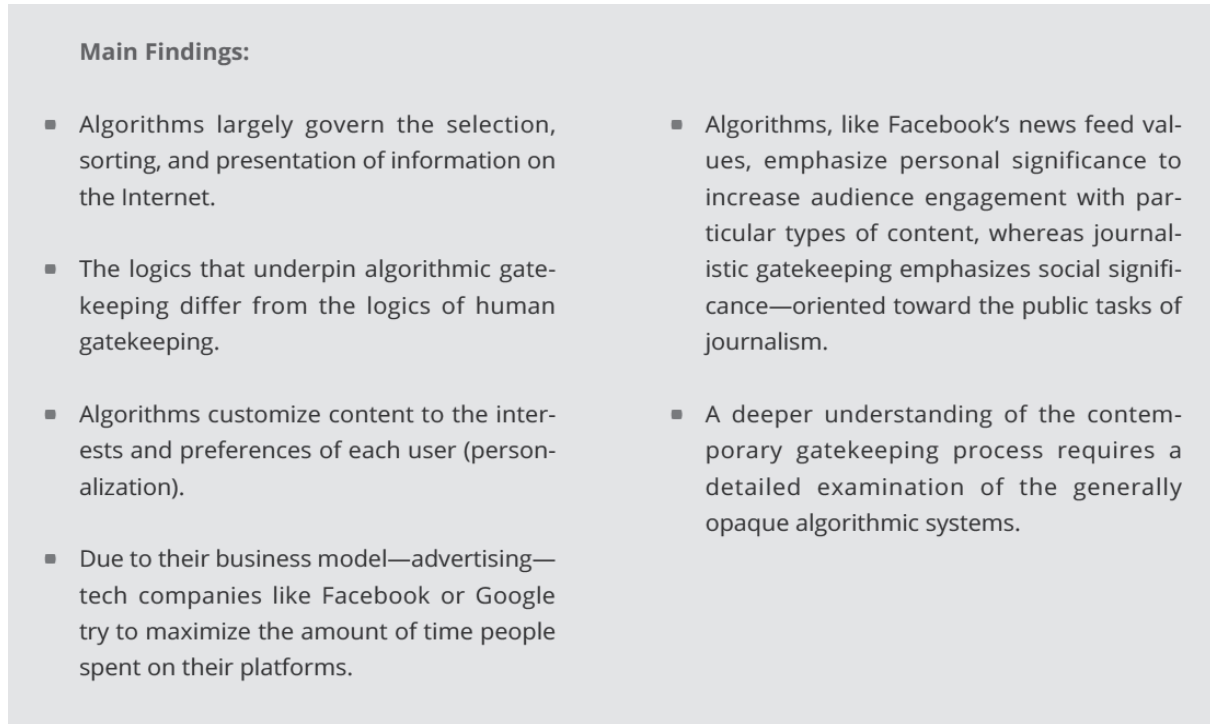


Figure 25: Main findings of Algorithmic Information Filtering.

Source: [2-60, p. 13]

Laura Chinchilla, former president of Costa Rica and chair of the Kofi Annan Commission on Elections and Democracy in the Digital Age, summed up the not-so-simple truths about digital social media platforms:

[...] On the one hand, digital technologies have played a vital role in providing free access to government data and information; encouraging citizen participation in public decision making; introducing new voices to the public debate; fostering the transparency and scrutiny of administrative actions; knitting global advocacies together on issues affecting human rights, the rule of law and democracy; and mobilizing new actors eager to find alternative avenues for political participation. The Arab Spring almost a decade ago, the pro-democracy protests in Hong Kong this summer and the toppling of Puerto Rico's governor in July are only a few examples.

On the other, the alarming number of episodes involving the use of social media platforms to manipulate elections and public debates, as well as the surge of extremist groups using the internet to incite hatred and violence, clearly warns us that the adverse relationship between those platforms and democracy is no longer just anecdotal.

Fake news is as old as news, and hate speech is as old as speech. But the digital age has provided a ripe environment for the virulent reproduction and visibility of both. To be clear, the promise of the betterment of the human condition held by new technologies is beyond question. But the risks have become just as apparent. [...] [2-70]

In summary, social media networks in particular can contribute to the polarization and radicalization of their users by reflecting a distorted picture of news as well as opinions and by reinforcing cognitive biases. Moreover, fake news spreads much faster than true news. The rapid spread was confirmed by an MIT study, whose results showed that *“the top 1% of false news cascades diffused to between 1000 and 100,000 people, whereas the truth rarely diffused to more than 1000 people”* [2-71].

References Chapter 2

- [2-1] UNITED NATIONS, Universal Declaration of Human Rights. Available at <https://www.un.org/en/about-us/universal-declaration-of-human-rights> (last accessed 06 January 2022).
- [2-2] EUROPEAN COURT OF HUMAN RIGHTS, Council of Europe's work on Hate Speech. Available at <https://www.coe.int/en/web/committee-on-combatting-hate-speech/council-of-europe-work-on-hate-speech> (last accessed 13 October 2021).
- [2-3] EUROPEAN COURT OF HUMAN RIGHTS, European Convention on Human Rights, Strasbourg 2013. Available at https://www.echr.coe.int/documents/convention_eng.pdf (last accessed 01 December 2021).
- [2-4] DEUTSCHER BUNDESTAG, Fake-News, Definition und Rechtslage, 2017. Available at <https://www.bundestag.de/resource/blob/502158/99feb7f3b7fd1721ab4ea631d8779247/wd-10-003-17-pdf-data.pdf> (last accessed 05 December 2021).
- [2-5] MACAVANEY, S., HAO-REN, Y., YANG, E., RUSSELL, K., GOHARIAN, N., FRIEDER, O., Hate speech detection: Challenges and solutions, in: PLoS ONE 14(8) (2019). Available at <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0221152> (last accessed 20 January 2022).
- [2-6] CARLSON, C. R., Hate Speech, Cambridge, Massachusetts: The MIT Press, 2021. Available at <https://doi.org/10.7551/mitpress/12491.003.0001> (last accessed 20 January 2022).
- [2-7] EUROPEAN COURT OF HUMAN RIGHTS, Hate speech, 2020. Available at https://www.echr.coe.int/documents/fs_hate_speech_eng.pdf (last accessed 18 October 2021).
- [2-8] EUROPEAN COURT OF HUMAN RIGHTS, Guide on Article 8 of the European Convention on Human Rights, 2021. Available at https://www.echr.coe.int/documents/guide_art_8_eng.pdf (last accessed 01 December 2021).
- [2-9] COUNCIL OF EUROPE, Digital Resistance, 2020. Available at <https://rm.coe.int/digital-resistance-handbook-21012021/1680a1269d> (last accessed 27 September 2021).
- [2-10] MAZZEO, V., RAPISARDA, A., GIUFFRIDA, G., Detection of Fake News on COVID-19 on Web Search Engines, 2021. Available at <https://www.frontiersin.org/article/10.3389/fpsy.2021.685730> (last accessed 20 January 2022).
- [2-11] EUROPEAN COMMISSION AGAINST RACISM AND INTOLERANCE, ECRI General Policy Recommendation No.15 on Combating Hate Speech: key points, 2016. Available at <https://rm.coe.int/16805d59ee> (last accessed 20 January 2022).
- [2-12] LAAKSONEN, S.-M., HAAPOJA, J., KINNUNEN, T., NILIMARKKA, M., PÖYHTÄRI, R., The Datafication of Hate: Expectations and Challenges in Automated Hate Speech Monitoring, 2020. Available at <https://www.frontiersin.org/article/10.3389/fdata.2020.00003> (last accessed 20 January 2022).
- [2-13] PATTON, D. U., FREY, W. R., MCGREGOR, K. A., LEE, F.-T.; MCKEOWN, K., MOSS, E., Contextual Analysis of Social Media: The Promise and Challenge of Eliciting Context in Social Media Posts with Natural Language Processing, in: Proceedings of the 2020

- AAAI/ACM Conference on AI, Ethics, and Society (AIES '20), New York 2020. Available at <https://doi.org/10.1145/3375627.3375841> (last accessed 20 January 2022).
- [2-14] COUNCIL OF EUROPE, Compilation of Venice: Commission opinions and reports concerning freedom of expression and media, in: European Commission for democracy through law, Strasbourg 2020. Available at [https://www.venice.coe.int/webforms/documents/?pdf=CDL-PI\(2020\)008-e](https://www.venice.coe.int/webforms/documents/?pdf=CDL-PI(2020)008-e) (last accessed 13 October 2021).
- [2-15] COUNCIL OF EUROPE, Information Disorder: Toward an interdisciplinary framework for research and policy making, in: Council of Europe report DGI (2017)09, Strasbourg 2017. Available at [0900001680766412 \(coe.int\)](https://www.coe.int/t/0900001680766412) (last accessed 13 October 2021).
- [2-16] MEISTER, S., Isolation and Propaganda: the roots and instruments of Russia's disinformation campaign, in: Transatlantic Academy Paper Series 2015-16 No.6, Washington DC 2016. Available at https://dgap.org/system/files/article_pdfs/meister_isolationpropoganda_apr16_web_1.pdf (last accessed 11 December 2021).
- [2-17] DOUBLET, Y-M., in: Disinformation and electoral campaigns, Council of Europe, Strasbourg 2019. Available at <https://book.coe.int/fr/droit-international/7985-disinformation-and-electoral-campaigns.html> (last accessed 20 October 2021).
- [2-18] TURČILO, L., OBRENOVIĆ, M., Fehlinformationen, Desinformationen, Malinformationen: Ursachen, Entwicklungen und ihr Einfluss auf die Demokratie, in: Heinrich Böll Stiftung, Demokratie im Fokus #3, 2020. Available at https://www.boell.de/sites/default/files/2020-08/200825_E-Paper3_DE.pdf (last accessed 03 November 2021).
- [2-19] AKERLOF, G.A., The Market for "Lemons": Quality Uncertainty and the Market Mechanism, in: The Quarterly Journal of Economics, Vol. 84 No.3, 1970.
- [2-20] DR. STAHEL, L., in: Status quo und Massnahmen zu rassistischer Hassrede im Internet: Übersicht und Empfehlungen, Soziologisches Institut, Universität Zürich, 2020. Available at https://www.edi.admin.ch/dam/edi/de/dokumente/FRB/Neue%20Website%20FRB/T%C3%A4tigkeitsfelder/Medien_Internet/Stahel_2020_Status%20quo%20und%20Massnahmen%20zu%20rassistischer%20Hassrede.pdf.download.pdf/Stahel_2020_Status%20quo%20und%20Massnahmen%20zu%20rassistischer%20Hassrede.pdf (last accessed 09 November 2021).
- [2-21] GABIELKOV, M., RAMACHANDRAN, A., CHAINTREAU, A., LEGOUT, A. (2016). Social Clicks: What and Who Gets Read on Twitter? ACM SIGMETRICS Perform. Eval. Rev. 44, 1 (June 2016), pp. 179–192, <https://doi.org/10.1145/2964791.2901462>.
- [2-22] SHAO, C., HUI, P.-M., WANG, L., JIANG, X., FLAMMINI, A., MENCZER, F., et al. (2018). Anatomy of an online misinformation network. PLoS ONE 13(4): e0196087, <https://doi.org/10.1371/journal.pone.0196087>.
- [2-23] SILVERMAN, C., ALEXANDER, L. How Teens In The Balkans Are Duping Trump Supporters With Fake News. BuzzFeedNews, November 3, 2016 (online), <https://www.buzzfeednews.com/article/craigsilverman/how-macedonia-became-a-global-hub-for-pro-trump-misinfo#.nfGBdzv3rN> (last accessed November 24, 2021).

- [2-24] KING, G., PAN, J., & ROBERTS, M. (2017). How the Chinese Government Fabricates Social Media Posts for Strategic Distraction, Not Engaged Argument. *American Political Science Review*, 111(3), pp. 484-501, doi:10.1017/S0003055417000144.
- [2-25] PENNYCOOK, G., RAND, D.G. (2019). Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition*, Volume 188, 2019, pp. 39-50, <https://doi.org/10.1016/j.cognition.2018.06.011>.
- [2-26] WEIR, K. Why we fall for fake news: Hijacked thinking or laziness? *American Psychological Association*, 11.02.2020 (online), <https://www.apa.org/news/apa/2020/fake-news> (last accessed December 4, 2021).
- [2-27] ROSS, R. M., RAND, D. G., PENNYCOOK, G. (2019). Beyond “fake news”: Analytic thinking and the detection of false and hyperpartisan news headlines. *PsyArXiv*, 13 November 2019, <https://doi.org/10.31234/osf.io/cgsx6>.
- [2-28] BAGO, B., RAND, D. G., PENNYCOOK, G. (2020). Fake news, fast and slow: Deliberation reduces belief in false (but not true) news headlines. *Journal of Experimental Psychology: General*, 149(8), pp. 1608–1613, <https://doi.org/10.1037/xge0000729>.
- [2-29] BRONSTEIN, M.V., PENNYCOOK, G., BEAR, A., RAND, D.G., CANNON, T.D. (2019). Belief in Fake News is Associated with Delusionality, Dogmatism, Religious Fundamentalism, and Reduced Analytic Thinking. *Journal of Applied Research in Memory and Cognition*, Volume 8, Issue 1, 2019, pp. 108-117, <https://doi.org/10.1016/j.jarmac.2018.09.005>.
- [2-30] PENNYCOOK, G., RAND, D.G. (2019). Who falls for fake news? The roles of bullshit receptivity, overclaiming, familiarity, and analytic thinking. *Journal of Personality*. 2020, 88, pp. 185– 200, <https://doi.org/10.1111/jopy.12476>.
- [2-31] SUNDAR, S.S., KNOBLOCH-WESTERWICK, S., HASTALL, M.R. (2007). News cues: Information scent and cognitive heuristics. *Journal of the American Society for Information Science and Technology*, 58, pp. 366-378, <https://doi.org/10.1002/asi.20511>.
- [2-32] SCHMITT-BECK, R. (2015). Bandwagon Effect. In *The International Encyclopedia of Political Communication*, G. Mazzoleni (Ed.), <https://doi.org/10.1002/9781118541555.wbiepc015>.
- [2-33] LEVITAN, L.C., VERHULST, B. (2016). Conformity in Groups: The Effects of Others’ Views on Expressed Attitudes and Attitude Change. *Political Behavior*, 38, pp. 277–315, <https://doi.org/10.1007/s11109-015-9312-x>.
- [2-34] MALLINSON, D.J., HATEMI, P.K. (2018). The effects of information and social conformity on opinion change. *PLoS ONE* 13(5): e0196600, <https://doi.org/10.1371/journal.pone.0196600>.
- [2-35] INSKO, C.A., SMITH, R.H., ALICKE, M.D., WADE, J., TAYLOR, S. (1985). Conformity and Group Size: The Concern with Being Right and the Concern with Being Liked. *Personality and Social Psychology Bulletin*, 1985, 11(1), pp. 41-50, <https://doi.org/10.1177/0146167285111004>.
- [2-36] KISS, Á., SIMONOVITS, G. (2013). Identifying the bandwagon effect in two-round elections. *Public Choice*, 160, 2014, pp. 327–344, <https://doi.org/10.1007/s11127-013-0146-y>.

- [2-37] MORWITZ, V.G., PLUZINSKI, C. (1996). Do Polls Reflect Opinions or Do Opinions Reflect Polls? The Impact of Political Polling on Voters' Expectations, Preferences, and Behavior. *Journal of Consumer Research*, Volume 23, Issue 1, June 1996, pp. 53–67, <https://doi.org/10.1086/209466>.
- [2-38] BENECKE, O., DEYOUNG, S.E. (2019). Anti-Vaccine Decision-Making and Measles Resurgence in the United States. *Global Pediatric Health*. January 2019, <https://doi.org/10.1177/2333794X19862949>.
- [2-39] GUESS, A.M., NYHAN, B., REIFLER, J. (2018). Exposure to untrustworthy websites in the 2016 U.S. election. Available at <https://www.dartmouth.edu/~nyhan/fake-news-2016.pdf> (last accessed 04.12.2021).
- [2-40] KAHAN, D.M. (2017). Misconceptions, Misinformation, and the Logic of Identity-Protective Cognition. Cultural Cognition Project Working Paper Series No. 164, Yale Law School, Public Law Research Paper No. 605, Yale Law & Economics Research Paper No. 575, Available at SSRN: <https://ssrn.com/abstract=2973067> or <http://dx.doi.org/10.2139/ssrn.2973067>.
- [2-41] KAHAN, D.M. (2013). Ideology, motivated reasoning, and cognitive reflection. *Judgment and Decision Making*, Vol. 8, No. 4, July 2013, pp. 407-424. Available at <http://journal.sjdm.org/13/13313/jdm13313.pdf>.
- [2-42] VAN BAVEL, J.J., PEREIRA, A. (2018). The Partisan Brain: An Identity-Based Model of Political Belief. *Trends in Cognitive Sciences* 22 (3), pp. 213–224, <https://doi.org/10.1016/j.tics.2018.01.004>.
- [2-43] PENNYCOOK, G., RAND, D.G. (2021). The Psychology of Fake News. *Trends in Cognitive Sciences* 25 (5), pp. 388–402, <https://doi.org/10.1016/j.tics.2021.02.007>.
- [2-44] SYDELL, L. We Tracked Down A Fake-News Creator In The Suburbs. Here's What We Learned. NPR, November 23, 2016 (online), <https://www.npr.org/sections/alltechconsidered/2016/11/23/503146770/npr-finds-the-head-of-a-covert-fake-news-operation-in-the-suburbs?t=1638618145394> (last accessed December 4, 2021).
- [2-45] GABBATT, A. 'Incited by the president': politicians blame Trump for insurrection on Capitol Hill. *The Guardian*, January 7, 2021 (online), <https://www.theguardian.com/us-news/2021/jan/06/donald-trump-politicians-insurrection-capitol-hill>, (last accessed December 4, 2021).
- [2-46] U.S. DEPARTMENT OF STATE (2020). GEC Special Report: Pillars of Russia's Disinformation and Propaganda Ecosystem. Available at https://www.state.gov/wp-content/uploads/2020/08/Pillars-of-Russia's-Disinformation-and-Propaganda-Ecosystem_08-04-20.pdf.
- [2-47] THORSON, E. (2015). Belief Echoes: The Persistent Effects of Corrected Misinformation. *Political Communication*, 33 (3) 2016, pp. 460-480, <https://doi.org/10.1080/10584609.2015.1102187>.
- [2-48] PENNYCOOK, G., BEAR, A., COLLINS, E.T., RAND, D.G. (2020). The Implied Truth Effect: Attaching Warnings to a Subset of Fake News Headlines Increases Perceived Accuracy of Headlines Without Warnings. *Management Science* 66 (11), pp. 4944–4957, <https://doi.org/10.1287/mnsc.2019.3478>.

- [2-49] DEL VICARIO, M., BESSI, A., ZOLLO, F., PETRONI, F., SCALA, A., CALDARELLI, G. et al. (2016). The spreading of misinformation online. *Proceedings of the National Academy of Sciences of the United States of America* 113 (3), pp. 554–559, <https://doi.org/10.1073/pnas.1517441113>.
- [2-50] BONFADELLI, H. *Medien und Gesellschaft im Wandel*. Bundeszentrale für politische Bildung, December 9, 2016 (online), <https://www.bpb.de/gesellschaft/medien-und-sport/medienpolitik/236435/medien-und-gesellschaft-im-wandel> (last accessed December 29, 2021).
- [2-51] LUHMANN, N. (1996). *Die Funktion der Massenmedien. Die Realität der Massenmedien*, VS Verlag für Sozialwissenschaften, Wiesbaden, https://doi.org/10.1007/978-3-663-01103-3_13.
- [2-52] SCHIFFER, S. *Der Islam in deutschen Medien*. Bundeszentrale für politische Bildung, May 10, 2005 (online), <https://www.bpb.de/apuz/29060/der-islam-in-deutschen-medien> (last accessed December 29, 2021).
- [2-53] JARREN, O., KÜNZLER, M., PUPPIS, M. (2012). *Medienwandel oder Medienkrise? Folgen für Medienstrukturen und ihre Erforschung*. Nomos Verlagsgesellschaft, pp. 9–25, <https://doi.org/10.5771/9783845236735-9>.
- [2-54] HOFFMANN, J., RAUPP, J. (2006). *Politische Personalisierung*. Pub 51, pp. 456–478, <https://doi.org/10.1007/s11616-006-0240-y>.
- [2-55] BERNHARD, U., SCHARF, W. (2008). „Infotainment“ in der Presse. Pub 53, pp. 231–250, <https://doi.org/10.1007/s11616-008-0077-7>.
- [2-56] BRÜGGEMANN, M., ENGESSER, S. (2017). Beyond false balance: How interpretive journalism shapes media coverage of climate change. *Global Environmental Change* 42, pp. 58–67, <https://doi.org/10.1016/j.gloenvcha.2016.11.004>.
- [2-57] HORSTKÖTTER, N., MÜLLER, U., OMMEN, O., PLATTE, A., RECKENDREES, B., STANDER, V., LANG, P., THAISS, H. (2017). *Einstellungen, Wissen und Verhalten von Erwachsenen und Eltern gegenüber Impfungen – Ergebnisse der Repräsentativbefragung 2016 zum Infektionsschutz. BZgA-Forschungsbericht*. Köln: Bundeszentrale für gesundheitliche Aufklärung. Available at https://www.bzga.de/fileadmin/user_upload/PDF/studien/infektionsschutzstudie_2016--f4f414f596989cf814a77a03d45df8a1.pdf (last accessed January 23, 2022).
- [2-58] DIXON, G. N., CLARKE, C. E. (2013). Heightening Uncertainty Around Certain Science: Media Coverage, False Balance, and the Autism-Vaccine Controversy. *Science Communication*, 35 (3), pp. 358–382. <https://doi.org/10.1177/1075547012458290>.
- [2-59] BETSCH, C., SCHMID, P., KORN, L. et al. (2019). *Impfverhalten psychologisch erklären, messen und verändern*. *Bundesgesundheitsblatt* 62, pp. 400–409, <https://doi.org/10.1007/s00103-019-02900-6>, also available at https://www.rki.de/DE/Content/Kommissionen/Bundesgesundheitsblatt/Downloads/2019_04_Betsch.pdf?__blob=publicationFile (last accessed January 23, 2022).
- [2-60] STARK, B., STEGMANN, D., MAGIN, M., JÜRGENS, P. (2020). *Are Algorithms a Threat to Democracy? The Rise of Intermediaries: A Challenge for Public Discourse*. AW

- AlgorithmWatch. Available at <https://algorithmwatch.org/en/wp-content/uploads/2020/05/Governing-Platforms-communications-study-Stark-May-2020-AlgorithmWatch.pdf> (last accessed January 23, 2022).
- [2-61] ROSS, L., LEPPER, M. R., HUBBARD, M. (1975). Perseverance in self-perception and social perception: Biased attributional processes in the debriefing paradigm. *Journal of Personality and Social Psychology*, 32 (5), pp. 880–892, <https://doi.org/10.1037/0022-3514.32.5.880>.
- [2-62] NICKERSON, R. S. (1998). Confirmation Bias: A Ubiquitous Phenomenon in Many Guises. *Review of General Psychology*, 2 (2), pp. 175–220, <https://doi.org/10.1037/1089-2680.2.2.175>.
- [2-63] SAXENA, R. The social media “echo chamber” is real. *ARS Technica*, March 13, 2017 (online), <https://arstechnica.com/science/2017/03/the-social-media-echo-chamber-is-real/> (last accessed January 04, 2022).
- [2-64] CIAMPAGLIA, G.L., MENCZER, F. Misinformation and biases infect social media, both intentionally and accidentally. *The Conversation*, June 20, 2018 (updated January 10, 2019) (online), <https://theconversation.com/misinformation-and-biases-infect-social-media-both-intentionally-and-accidentally-97148> (last accessed January 04, 2022).
- [2-65] FERRARA, E., VAROL, O., DAVIS, C., MENCZER, F., FLAMMINI, A. (2016). The Rise of Social Bots. *Communications of the ACM*, July 2016, Vol. 59 No. 7, pp. 96-104, 10.1145/2818717. Available at <https://cacm.acm.org/magazines/2016/7/204021-the-rise-of-social-bots/fulltext> (last accessed January 04, 2022).
- [2-66] VAROL, O., FERRARA, E., DAVIS, C.A., MENCZER, F., FLAMMINI, A. (2017). Online Human-Bot Interactions: Detection, Estimation, and Characterization. *AAAI Publications*, Eleventh International AAAI Conference on Web and Social Media. Available at <https://aaai.org/ocs/index.php/ICWSM/ICWSM17/paper/view/15587> (last accessed January 25, 2022).
- [2-67] ADAMS, T. Facebook’s week of shame: the Cambridge Analytica fallout. *The Guardian*, March 24, 2018 (online), <https://www.theguardian.com/technology/2018/mar/24/facebook-week-of-shame-data-breach-observer-revelations-zuckerberg-silence> (last accessed November 09, 2021).
- [2-68] HERDER, J. (2018). *Regieren Algorithmen? Über den sanften Einfluss algorithmischer Modelle. (Un)berechenbar? Algorithmen und Automatisierung in Staat und Gesellschaft*, pp. 179-203, ISBN: 978-3-9818892-5-3, <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-57547-2>.
- [2-69] GILLESPIE, T. (2016). Algorithm. In B. Peters (Ed.), *Digital Keywords: A Vocabulary of Information Society and Culture*, pp. 18–30. Princeton University Press, <https://doi.org/10.2307/j.ctvct0023.6>.
- [2-70] CHINCHILLA, L. Post-Truth Politics Afflicts the Global South, Too. *The New York Times*, October 15, 2019 (online), <https://www.nytimes.com/2019/10/15/opinion/politics-global-south.html>, also <https://www.kofiannanfoundation.org/supporting-democracy-and-elections-with-integrity/annan-commission/post-truth-politics-afflicts-the-global-south-too/> (last accessed December 09, 2021).
- [2-71] VOSOUGHI, S., ROY, D., ARAL, S. (2018). The spread of true and false news online. *Science* (New York, N.Y.) 359 (6380), p. 1146–1151, <https://www.science.org/doi/10.1126/science.aap9559>.