2.8
COUNTERING CONSPIRACY THEORIES AND MISINFORMATION

Péter Krekó

Introduction

There are many research studies on the functions and background motives underlying belief in conspiracy theories (Douglas et al. 2017) and also the negative consequences of conspiracy theories (Douglas et al. 2015; see also Chapter 2.7 in this volume). However, we so far only have a limited knowledge of the most practical implications of conspiracy theories: How they can be changed, debunked and modified. This chapter tries to systemically overview and summarise the most important research so far concerning the possibilities of changing conspiracy beliefs via targeted interventions.

I do not take it for granted that there is a consensus over the need for interventions. In the beginning of the chapter, we take a look at the epistemological, moral and democratic arguments on whether, and when, we need to use interventions to reduce conspiracy beliefs. Then we briefly overview some psychological obstacles in the way of interventions. In the next section, we propose a matrix as a theoretical framework for categorising the possible interventions and overview the available academic literature as well as some practical experiences concerning efficient ways of reducing conspiracy beliefs. In the final section, we identify a broader avenue for future research.

To debunk or not to debunk? That is the question

The majority of the psychological research on conspiracy theories is based on the facts that most conspiracy theories are possibly harmful (van Prooijen, van Lange 2014; Douglas et al. 2015; van Prooijen 2018; see also Chapter 2.7 in this volume). However, before we jump into the practicalities of interventions and debunking, it is worth elaborating on the question of why and when interventions against conspiracy theories are needed and justified. We will consider three arguments why conspiracy theories may not always need interventions: the epistemic, the moral and the democratic argument. Then, based on these considerations, we identify the theories we think are the most important to get debunked.

1. The epistemic argument. As some philosophers argue, conspiracy theories can be, in some cases, accurate descriptions of events in the world since history is full of conspiracies and conspiratorial assumptions (e.g. Pigden 1995). Also, conspiracy theories can sometimes be the
Countering conspiracy theories

official explanation of events, as in the case of the Iraq War, where the *casus belli* was the alleged existence of weapons of mass destruction in Iraq. It proved to be false – raising a new suspicion of conspiratorial intentions of political leaders who might have been aware of their non-existence, but manipulated the public to reach their political goals (Coady 2006). Therefore, a priori rebuttal of conspiracy theories can be problematic. Also, a lack of convincing evidence is not necessary proof for their non-existence. ‘Whatever we might think about conspiracy theories generally, there is no prima facie case for a scepticism of conspiracy theories based purely on their use of evidence’ (Dentith 2017: 1) Also, in some cases, we have some good reasons to question the knowledge that comes from traditional sources, especially when this ‘knowledge’ is still not rock-solid, and/or there is contradicting evidence on the field. As Uscinski (2018: 5) put it: ‘Our knowledge generating institutions are not—or should they be above suspicion or reproach’.

2. The moral argument. Some scholars who are more ‘empathic’ towards conspiracy theories argue that conspiracy theories are in fact beneficial, as they can foster more transparency via challenging the established social hierarchy as narratives of non-privileged groups (Fenster 1999; Clarke 2002). Also, sometimes believers of conspiracy theories can embrace democratic principles, as it was found among the supporters of the 9/11 conspiracy theories (Swami et al. 2010). Conspiracy theories, according to this line of argumentation, can serve as alarm bells and early warning systems (Uscinski 2018) and articulate the concerns of the ‘losers’ (Uscinski et al. 2011).

3. Democratic argument: Freedom of speech and pluralism. The third line of arguments is based on the notion that conspiracy theories are a legitimate part of the democratic discussion and deserve defence as a form of free speech as ‘a crucial part of the marketplace of ideas’ (Uscinski 2018: 6.). According to this perspective, debunking and regulation might be problematic as a form of political suppression.

Of course, while some of these arguments contain valid points, interventions against conspiracy theories are still important in most of the cases. At the same time, even without a general taxonomy, we can identify a sub-group of conspiracy theories that need to be challenged, based on the considerations above. The features of conspiracy theories that require urgent interventions share three characteristics: harmfulness, low plausibility and high popularity. However, not all of the three conditions should be fulfilled in order for an intervention to be justified. For example, in the case of ideological conspiracy theories of terrorist groups in the Western world, they are definitely not popular, but at the same time can be extremely harmful. Also, theories that are non-plausible but popular should be debunked (even if not prohibited) if they are not directly harmful.

1. Plausibility. While conspiracies sometimes occur, most conspiracy theories are not the most plausible explanations of events (e.g. Aaronovitch 2011). Not every conspiracy theory is necessarily false, but most of them contain logical pitfalls and weaknesses. While our capacity is limited when trying to separate the true conspiracy theories from the ones that are not true, several conspiracy theories fail to pass the most basic logical tests. There are logical tools to examine the veracity of conspiracy theories on an a priori basis or based on the available information (for an overview, see Coady 2006). Probably the most important logical fallacy behind conspiracy theories is that they assume that a massive conspiracy can take place and remain a secret at the same time. Grimes (2016) made an estimation of how many people would need to have been involved in the plot to make some conspiracies happen and found extremely high figures: U.S. moon landings were a hoax (411,000 people); climate change is a fraud (405,000 people); unsafe vaccinations are being covered up (22,000); the cure for cancer is being suppressed by the world’s leading pharmaceutical firms (714,000 people).
2. Impact. Some conspiracy theories are harmless in themselves. For example, if one believes in lizard people, U.F.O.s, flat earth, the moon landing theory is fake – these beliefs (if not systematically connected to other conspiracy beliefs – which is often the case) are not necessarily causing any harm or manifesting in some problematic behaviour. At the same time, we can say that believing in these theories can dramatically erode trust in public institutions (lizard men, moon landing hoax) or spread distrust in scientists (flat earth), which can be an important reason to counterargue and debunk them (but not to prohibit them). However, many well-known and relevant conspiracy theories are directly harmful – not only for the individual but for broader society, as has been found in many studies (for a summary, see Douglas et al. 2015).

Political conspiracy theories can lead to alienation, cynicism and political disengagement (Butler et al. 1995). Exposure to conspiracy theories decreases trust in governmental institutions, even if the conspiracy theories are unrelated to those institutions (Einstein, Glick 2015). Conspiracy theories not only can cause disenchantment with politicians, but also with scientists and scientific institutions (Jolley, Douglas 2014a). Scientific findings such as vaccination, that can prevent diseases such as measles, are even questioned by some conspiracy theories, and exposure to these conspiracy theories can lead to lower vaccination intentions (Jolley, Douglas 2014a). More generally, medical conspiracy theories can lead to irrational health behaviour that shortens or terminates lives (Oliver, Wood 2014). Conspiracy theories about the H.I.V. virus (that it was created in governmental laboratories) can lead to reduced condom use as both are associated with the same belief of genocidal intentions against blacks by the U.S. government (Bogart, Thorburn 2005), therefore raising the chance of H.I.V. infection and unintended pregnancy. Also, conspiracy theories about climate change, such as that global warming is a hoax, can discourage steps to reduce one’s carbon footprint (Jolley, Douglas 2014b; Van Der Linden et al. 2015). There may also be a more general ‘conspiracy effect’ such that exposure to conspiracy theories decreases pro-social behaviour and acceptance of scientific findings (Van Der Linden 2015). Conspiracy theories against Jews are the most important predictor of negative behavioural intentions towards them (Bilewicz et al. 2013). Conspiracy theories against outgroups can also serve as ‘radicalization multipliers’ (Bartlett, Miller 2010) and encouraging ‘apocalyptic aggression’ towards outgroups that seem to pose a danger (Berlet 2009).

While conspiracy theories are often stigmatised knowledge (Barkun 2003), in many cases this is justified. Popular conspiracy theorists such as Alex Jones provide very few arguments regarding when these theories should be defended, as they are epistemically wrong, morally malevolent and dangerous for democracies.

3. Popularity. Any countermeasures towards conspiracy theories should mainly target the popular theories that are shared by a significant portion of society, for two simple pragmatic reasons. First, popular theories are more likely to lead to negative consequences (see above). Second, refuting less well-known theories can have a countereffect: Making the previously unknown theories more known and popular. The popularity of conspiracy theories is sometimes the result of their political manufacturing. Conspiracy theories these days are not necessarily the ‘theories of losers’, but official or semi-official narratives of events that are spread by non-democratic actors. We can see current examples in Turkey, Russia and North Korea, but also Poland and Hungary. Conspiracy theories are spread by powerful actors such as Russia, using all the resources of the state (e.g. Yablokov 2015, 2018). Conspiracy theories exploiting a victim identity are disguised to be the tools of the weak and oppressed and are often dangerous tools in the hands of the strong and powerful. Solidarity with the powerless therefore should not translate to solidarity with conspiracy theorists.

Based on the summary above, when we talk about counterstrategies and debunking, we will mostly focus only on the theories that fit the three conditions set above: Non-plausible, harmful
and popular at the same time. As we see from the list above, this is still a very broad category within conspiracy theories. Many conspiracy theories that are popular and not so plausible are harmful as well. For example, ‘global warming is a hoax’ theories that discourage efforts to save the environment as well as decreasing trust in science, are shared by 37 per cent of the U.S. electorate, and more than 60 per cent of the Republican voters in the U.S. (Jensen 2013).

The difficulties of debunking conspiracy theories

It is easier to spread conspiracy theories than to refute them. Conspiracy theories have a strong functional basis: They can serve epistemic, existential, social motives at the same time (Douglas et al. 2017). Even if conspiracy theories are sometimes self-defeating forms of motivated cognition, their refutation is often extremely difficult. We can mention at least seven important psychological obstacles when we would like to debunk, or change, false beliefs, to take these factors into considerations when tailoring counterstrategies.

1. The resistance of conspiracy theories to factual information. Conspiracy theories are often non-falsifiable (Keeley 1999; Uscinski, Parent 2014), as they are referring to invisible, secret, dark dealings behind the curtain. Therefore, the lack of visibility of the conspiracy in a lot of cases is used as an argument for its success and its existence. Furthermore, refuting attempts are often incorporated in a ‘metaconspiracy theory’: The belief that people and institutions who try to debunk conspiracy theories may, themselves, be part of the conspiracy (Lewandowsky et al. 2015, 2013). Disconfirming evidence (e.g. scientific findings) can therefore be regarded as products of a conspiracy. Given that traditional knowledge-producing institutions are perceived as part of the ‘establishment’, their arguments are particularly easy targets in the eyes of conspiracy theorists. But this principle is not universal: the resistance of factual information is mainly an obstacle for changing the minds of the ones who already have very strong conspiracy beliefs: the ‘true believers’.

2. Lingering effect of misinformation: ‘mud sticks’. As Ecker et al. (2017) found, even after careful and repeated counter-argumentation, the original information still leaves traces on people’s attitudes, as the cognitive representation does not disappear from memory. There is no such thing as perfectly refuted misinformation. This is a strong argument for interventions that aim to prevent exposure to a conspiracy theory.

3. The sleeper effect. As was found in early persuasion research and then confirmed by more recent metanalytic overviews (Kumkale, Albarracín 2004), if the information comes through with some discounting cue (e.g. the messenger is a highly unreliable source), it will still have an impact on the attitudes of the recipient. Furthermore, as time passes, the discounting cue, that is the ‘tag’ of lack of credibility, can gradually disappear from the original information. Therefore, the information prevails without the warning sign of non-credibility. It can be especially important in the case of fake news and conspiracy theories that mainly originate from highly unreliable sources. Even if the person is aware of the tainted nature of the source, the fake information can prevail in memory while the information of the source fades away. The cure for this problem could be to focus more on the content of the conspiracy theory than the source during the interventions, or to repeat the disconfirming information.

4. Familiarity Backfire Effect. In order to debunk a belief, one has to repeatedly mention it. However, this can have an unintended effect: The debunking effort itself can strengthen the belief by increasing its familiarity. Or, in the case that the belief was unknown, the debunking can spread the belief (Cook, Lewandowsky 2011). In the case of conspiracy theories, this question can be particularly relevant given that some conspiracy theories are rather unknown. To avoid the trap of unintentionally spreading conspiracy theories, we should therefore very
Péter Krekó

carefully evaluate each conspiracy theory to determine if it deserves refutation or not. If a theory is not widespread (see popularity as a precondition above), or because of its weak presence in the public is therefore probably not highly accessible, debunking can do more harm than good.

5. Reactance. Individuals tend to reject persuasion efforts if they feel that the messenger wants to limit their personal freedom of opinion or limit their behavioural alternatives (Miron, Brehm 2006). The result can be an even stronger embrace of this particular belief to prove that the individual makes sovereign decisions. This danger is also particularly apparent when it comes to conspiracy theories, as it is difficult not to formulate the debunking message as ‘lecturing’ that comes from the position of a kind of intellectual authority. Too intense, propagandistic counter messaging can lead to stronger reactance.

6. Identity and motivated reasoning. It is difficult to refute group-based beliefs on the level of the individual. This is what Cook and Lewandowsky (2011) call the Worldview Backfire Effect. While most of the debunking and counter messaging efforts focus on refuting conspiracy theories as a form of ‘crippled epistemology’ and assume that providing adequate information is the cure, this ‘information deficit model’ is often totally insufficient, as conspiracy theories can be deeply rooted in group identities and a wider set of beliefs. Given that conspiracy theories are forms of ‘motivated collective cognition’ (Krekó 2015), they can serve important advantages for the ingroup, such as raising collective self-esteem, putting the blame on external actors and strengthening ingroup identity via self-victimisation. Also, more intense communication of a position can lead to a boomerang-effect via further polarising opinions about an issue, and just strengthening the beliefs of conspiracy theorists – as it was found in the case of climate change (Hart, Nisbet 2012).

7. True believers and vulnerable groups. Connected to the previous point, the conspiracy beliefs of ‘true believers’ (the ones for whom conspiracy beliefs constitute a strong, central part of their belief system) are more difficult to change. Generally, more central, extreme and polarised attitudes are resistant to change (Banas, Miller 2013). Also, conspiracy theories are frequently strongly connected to extreme political positions that are strong and stable (van Prooijen et al. 2015). Furthermore, some social groups are traditionally more receptive to conspiracy theories than others, notably low status groups, minority groups and groups with lower levels of education (for a summary, see Chapter 2.6). These groups can be logical targets of interventions, especially immunisation – but the problem is, they are more difficult to reach via institutions. The problem of ‘true believers’ and vulnerable groups can be especially acute when dealing with demand-side solutions (see below).

The typology of counterstrategies

While there is abundant research on what induces conspiracy theories, the literature on how to reduce them is thinner. Given this limitation, in the section below I try to give a theoretical framework for the possible tools for debunking conspiracy theories, and also highlight some experiences beyond the strictly academic literature. According to this simple framework, every intervention aimed to counter conspiracy theories can be put on a two-dimensional matrix. The first dimension is temporal: does the intervention take place before the person meets with the conspiracy theory (prevention), or does it target a belief that the person has already encountered (harm reduction)? The second dimension deals with the target of the intervention: is it the messenger/source of the information (supply side) or the recipients of the intervention (demand side)?

This matrix gives us four possible options: (1) the intervention takes place before the spread of the message, and targets the supply side (pre-emptive strike); (2) it takes place before the
Countering conspiracy theories

dissemination, but targets the recipient (immunisation) (3); it takes place after the message was spread and targets the source, (striking back); or (4) it takes place after the dissemination but targets the recipient (healing).

In the remaining sections of this chapter, we go through these interventions one by one, evaluating them in the light of the academic literature and non-academic experiences so far.

Exposure to conspiracy theories is an important predictor behind such beliefs (Swami et al. 2010). Also, as discussed above, conspiracy theories are difficult to refute after the person was exposed to them. Furthermore, conspiracy theories have a hidden impact on the recipients’ attitudes and behaviour (Douglas, Sutton 2008). That is, they can influence people without them even knowing that they have been influenced. For these reasons, preventative interventions, when possible, are better than focusing on harm reduction. Therefore, we start our discussion with the preventive strategies.

Immunisation

‘Theories of conspiracy represent a permanent temptation for us all’ (Groh 1987: 2). Interventions focusing on immunisation aim to shield the audience with the necessary cognitive, emotional or motivational skills to resist this temptation. However, while it sounds like the ‘magic wand’ solution against conspiracy theories, it is not: People are often motivated to accept conspiracy theories and this is especially the case for some more vulnerable groups (see the ‘difficulties we face when debunking conspiracy theories’ above). Still, immunisation has been found to be successful in many cases. It can be general, aimed at reducing the general receptivity to conspiracy theories; or particular, focusing on specific beliefs.

Given that conspiracy thinking is often a result of a lack of logical thinking (e.g. Sunstein, Vermeule 2019), a general way of reducing the receptivity to conspiracy theories in advance can be the improvement of logical, analytic thinking. Swami et al. (2014) found that verbal fluency and cognitive fluency tasks eliciting analytic thinking reduced belief in conspiracy theories and conspiracist ideation.

Targeting the root causes of conspiracy theories can be another efficient approach for reducing conspiracy beliefs. While certain personality traits, social status or minority status are difficult or impossible to change, reducing feelings of uncertainty, mistrust, powerlessness and lack of control, that are found to be important factors that enhance conspiracy thinking, seems to be a possible way (Hamsher et al. 1968; Mirowsky, Ross 1983; Abalakina-Paap et al. 1999; Imhoff, Bruder 2014; Newheiser et al. 2011). While using interventions that reduce powerlessness and uncertainty on a social scale sound naïve and utopian, interventions in organisations might have a positive effect on trust and feelings of control that reduce conspiracy theorising (Kramer 1999; van Prooijen, de Vries 2016; Douglas, Leite 2017). To be more specific about the way of rebuilding feelings of control, we can think about reversing one important scientific finding. Specifically, many studies have found that it is easy to induce conspiracy thinking via recalling personal experiences where people failed to exercise control and therefore felt uncertain (Whitson, Galinsky 2008; Sullivan et al. 2010; van Prooijen, Acker 2015). Recalling experiences of successfully controlled events, on the other hand, might strengthen the feeling of self-efficacy and might contribute to reducing conspiracy ideation.

Also, a less self-evident specific way of immunisation against conspiracy theories can be the deliberate ignorance of false information to avoid its impact on our thinking (Hertwig, Engel 2016). It is a bit counterintuitive as our post-enlightenment ideals may suggest that knowing more is always better than knowing less. However, tactical avoidance of false information can
protect the person from the inevitable impact of blatant lies, false information and conspiracy theories. The problem is, though, that this is very difficult to do, given that some popular conspiracy theories in politics (such as climate scepticism, Pizzagate, 9/11) surround people in the ‘populist zeitgeist’ (Mudde 2004) they live in and people can only avoid meeting them if they close their minds to political information in general.

Immunisation can also be specific, targeting particular conspiracy beliefs. Preliminary warning seems to be an efficient solution in raising awareness about the possible misleading nature of the information that the individual is going to encounter. As Cook and Lewandowsky (2011) propose, any mention of false information should be preceded by warnings to notify the audience that the upcoming information is not true. Similarly, inoculation strategies that present weak arguments of the conspiracy theory in advance, in order to raise the attitudinal immune system of the person against such threats in the future, have been found efficient in the case of misinformation on climate change (Cook et al. 2017). Jolley and Douglas (2017) found inoculation an effective tool against anti-vaccination theories, underlining that the counterargument is efficient only when presented before meeting with misinformation. Banas and Miller (2013) also found that inoculation strategies can be effective in creating resistance to conspiracy theory propaganda, with fact-based (vs. logic-based) treatment being the most effective. However, this latter study also revealed that inoculation is a double-edged sword that can be a dangerous weapon in the hands of conspiracy theorists as well. The positive effect of inoculation can be significantly diminished if it is used against the inoculated arguments, and therefore it can defend the original conspiracy theory (called meta-inoculation).

**Pre-emptive strike**

Pre-emptive strikes against conspiracy theories aim to target the source of misinformation in order to reduce the chance that the target audience will encounter the undesirable piece of information. If this measure is successful, the piece of information will have no impact at all. The most obvious – but, at the same time, morally and technically most problematic – way of doing so is the removal of the harmful content or banning the source of misinformation from certain platforms. Especially after 2016 (the Brexit referendum and the U.S. presidential election campaign), social media companies came under increasing pressure to step up more harshly against disinformation, including conspiracy theories. Still, any specific strategy is controversial, and whilst some argue that social media companies are doing nothing, or not enough, against this challenge (e.g. Applebaum 2019), others argue that social media companies already poses an existential threat to freedom of speech (French 2019).

When it comes to the most popular social media platforms, Facebook’s community standards (Facebook 2019) reserve the right to remove ‘objectionable content’: Hate speech, content that ‘glorifies violence or celebrates the suffering or humiliation’ or is ‘cruel or insensitive’. Conspiracy theories that fall within these categories are often also targeted for removal. For example, in 2019, Facebook removed, among others, the profile of well-known conspiracy theorist Alex Jones, claiming that they remove sites that ‘promote or engage in violence and hate, regardless of ideology’ (Salinas 2018). Twitter made a similar decision, along with some other social media providers in 2016 (Salinas 2018).

Another typical reason for removal of sites that are spreading conspiracy theories can be malign foreign influence, and especially interference into elections. For example, in June 2019, Twitter announced that it banned almost 5000 accounts that it thought to be associated with foreign governments, claiming that Twitter has a responsibility to hamper ‘attempts to manipulate TWITTER to influence elections and other civic conversations by foreign or domestic...
Countering conspiracy theories

state-backed entities’ (Roth 2019). Following a similar logic, Twitter banned the advertisements of Russian propaganda channels RT and Sputnik that are spreading conspiracy theories on an industrial scale (see for example Yablokov 2018), referring to the assessment of the U.S. intelligence agencies that they aimed to influence the result of the 2016 U.S. presidential elections (Twitter Public Policy 2017). This is also an important step, as no advertisement can substantially reduce the ability of these accounts to reach the target audience.

Facebook’s new – and probably most fruitful – attempt is to ‘bury’ disinformation instead of ‘banning’ it – which means, significantly downplaying the chance of ‘low quality’ content appearing in one’s News Feed or searches. This occurs through changing the algorithms. Pages that contain malicious or deceptive advertisements can be significantly downgraded (Facebook Business 2017). Facebook, at the same time, have a targeted policy against anti-vaccination sites that are often spreading conspiracy theories about the malicious impact of the vaccinations (Bickert 2019). Fight against anti-vaccine misinformation contains the following specific steps: Reduction of ranking of groups and pages that spread misinformation about vaccinations in news feeds and searches; rejecting advertisements that spread misinformation about the vaccines, removing related targeting options and disabling accounts that violate these policies. Of course, such interventions are the targets of intense political, moral and legal discussions (see, for example, Wood and Ravel 2017). Also, it is not only the social media companies that are exposed to such controversies: The French ‘fake news law’ was passed in 2018 (Young 2018) as a reaction to the wave of disinformation during the 2017 French presidential election campaign (Ferrara 2017).

A possible tool for journalists can be the ignorance of the source of disinformation and avoid covering disinformation pieces (see, for example, Hellman and Wagnsson 2017). This is what the French press did during the last phase of the French presidential election campaign in 2017. When President Macron’s emails were leaked, the mainstream press refused to cover the content, and instead focused on the fact that hacking was used as a part of the influence operation, ignoring the content of the hacked material (Prier 2017, Lewandowsky 2019).

Based on the research above suggesting that conspiracy theories leave a trace on the psychological processes of the individual, these pre-emptive strike approaches, from a merely practical point of view, can be efficient. However, the broader assessment of impact of the measures aiming to ignore/exclude/tilt certain contents from the information environment are rather lacking. This is especially the case when it comes to conspiracy theories since the advocates of these theories can only feel assured that their theories about ‘global censorship’ of information are correct. Of course, most interventions do not focus on the ‘true believers’ but the general public, and less access to conspiracy theories can lead to less belief in them.

Healing

If the target audience has already encountered conspiracy theories, one possibility is to try to minimise their negative impact. The most obvious form of this behaviour is debunking the arguments of the conspiracy theories in the hope that it will discredit them in the eye of the observer. Orosz et al. (2016) found in experimental research that rational and ridiculing arguments were effective in reducing conspiracy beliefs (with modest effect sizes), whereas empathising with the targets of conspiracy theories had no effect. Individual differences played no role in the reduction of conspiracy beliefs, but the perceived intelligence and competence of the individual who conveyed the conspiracy belief-reduction information contributed to the success of reducing beliefs. Considering these results and previous studies focusing on the benevolent effects of analytic thinking (Swami et al. 2014), and fact-based inoculation (Banas, Miller 2013)
in reducing conspiracy beliefs, it can be assumed that uncovering arguments regarding the logical and factual inconsistencies of conspiracy beliefs can be an effective way to discredit them. This statement goes against the mainstream of the communication literature and the notion of ‘post-truth’ in which emotions trump all other factors. One possible explanation why rationality might have a bigger impact on shaping (sometimes irrational) beliefs than previously expected might be that, in the current communication environment, people are overloaded with emotional messages coming from advertisements, and political and social campaigns, and therefore rational and factual arguments have more informational value.

According to Cook and Lewandowsky (2011), efficient debunking requires three major elements. First of all, refutation must focus on the facts instead of the myth or false belief to avoid the misinformation becoming more familiar. Second, mentions of the false information should be preceded by warnings to notify the audience of the falseness of the upcoming information. Third, refutation should include an alternative explanation. Finally, to avoid the ‘overkill backfire effect’, that the refutation fails to persuade the audience because of too many counterarguments, less but strong arguments are usually more efficient than more but weaker arguments. Conspiracy theory-related metacognition can also be an efficient tool in diminishing the impact of such beliefs. Conspiracy exposure increases conspiracy beliefs and reduces trust, but asking about the beliefs that the recipient just heard facilitates additional thinking about the conspiracy claims and can soften or even reverse the effect of exposure on the conspiracy beliefs and their impact on trust.

One other possible solution is ‘cognitive infiltration’ (Sunstein, Vermeule 2009), that, according to its proponents, can be an efficient tool for breaking up the closed epistemic universe of ‘the hard core of extremists’. In this tool, government agents or their allies (acting either virtually or in real space, and either openly or anonymously) will undermine the crippled epistemology of believers by planting doubts about the theories and stylised facts that circulate within such groups, thereby introducing beneficial cognitive diversity.

(Sunstein, Vermeule 2009: 95)

While sowing the seeds of doubt can in itself bring beneficial effects, this idea had mixed reception at best. The fact that a former media advisor to President Obama proposed that the state interferes into private discussions served as a proof for the worst nightmares of conspiracy theorists, and inspired books as well (Griffin 2011). Also, of course, diversity of views is not enough in itself: In order for any messenger to challenge the attitudes of a group, they should affirm other beliefs of that individual or group or the values or position of the group, to gain more sympathy for any attitude change (Sherman, Cohen 2006; Feygina et al. 2010; Sunstein 2014; Wolsko et al. 2016).

**Striking back**

Interventions that target the source of information after the encounter with the message are rather rare, for the reason that this is not a particularly successful approach. Of course, measures of removing certain social media platforms because of their interference into the U.S. election are following this kind of ‘retaliatory’ logic, but we discussed them under the ‘pre-emptive strike’ solution because the real function of these measures is not to do harm for the source of the message after the damage is done, but to prevent further damage for other recipients. Discrediting the source of the information to do reputation damage fits in this logic, but according to the research on ‘sleeper effects’ (see above), it is not always highly efficient.
The logic of ‘tagging’ the information in an aim to undermine the credibility also follows the logic of ‘striking back’: The recipients become aware of the fact that this information was categorised as non-trustworthy by some players. Unfortunately, tagging items does not necessarily reduce their reach, and can rather trigger the opposite effect. Practical experience by Facebook that used this technique in cooperation with certified fact-checkers has shown that this tool can backfire and tagging certain articles as ‘disputed’ can only boost the interest towards them. As the author of misinformation told the Guardian: ‘A bunch of conservative groups grabbed this and said, “Hey, they are trying to silence this blog – share, share,” (...) With Facebook trying to throttle it and say, “Don’t share it,” it actually had the opposite effect.’ (Levin 2017). At the same time, tagging is not always unsuccessful in raising scepticism towards the information provided. It was found, for example, that tagging messages as advertisements can decrease their persuasive impact. While consumers rarely notice the promotion on the tweets, tweets promoted by a political party increased recognition of the advertisement and thus reduced behavioural intention, increased scepticism and negatively affected the source and also the attitudes. This effect, however, was not present for corporate brands (Boerman, Kruikemeier 2016).

Fact-checking is an extensively used tool to undermine epistemologically flawed information – a tool that can be used against conspiracy theories as well. Existing sites such as snopes.com, euvisinfo.eu, stopfake.org deal a lot with conspiracy theories. Fact-checking services can be used to filter and remove news that contains false information or to make consumers realise that what they have encountered is false. The efficiency of fact-checking sites is the subject of debates. Ecker et al. (2017) found that retractions (especially ones that explicitly repeated the misinformation) were efficient in reducing misinformation effects. At the same time, however, fact-checking efforts can have only an impact on the higher educated and the better informed, and trust in fact-checkers is not even on the political spectrum: Democrats tend to trust more in fact-checkers than Republicans (Nyhan, Reifler 2015). Also, not surprisingly, voters selectively share fact-checking messages that are beneficial for their own candidate and the opposing party’s candidate (Shin, Thorston, 2017), which seriously limits the opportunities of objective fact-checking. The examination of Twitter networks spreading low credibility information found that moving from the periphery to the core of the network, fact-checking nearly disappears. Again, this highlights some limits of its efficiency (Shao et al. 2018).

A different approach that can be highly successful is cutting the financial resources of conspiracy websites via warning the social media advertisement companies not to channel advertisement to these sites. Konspiratori.sk first published a list of conspiracy websites spreading harmful content (Čižik 2017). Webpages spreading fake news often generate income by selling advertising space, thus konspiratori.sk tries to make advertisers aware and help them to prevent their advertisements appearing on such websites.

Conclusion

This chapter provided an overview of the existing research and practical experiences of interventions against conspiracy theories. We found that theories that are popular, non-plausible and harmful are the ones that are obviously calling for some interventions. Intervention is difficult, and it has to overcome many obstacles.

The chapter proposed a matrix that, in our view, can be a useful but simple tool to categorise the possible ways of interventions using two dimensions: temporal (before or after the recipient met with the conspiracy theory), and the target (source – supply side vs. the recipient – demand side). It highlights four possible options: the intervention takes place before the spread of the
message, and targets the supply side (pre-emptive strike), or it takes place before the dissemination, but targets the recipient (immunisation); it takes place after the message was spread and targets the source (striking back), or takes place after the dissemination but targets the recipient (healing). Most of the research suggests that prevention is more useful than harm reduction, as conspiracy theories that have already spread are already having an impact on the recipient that is often difficult to neutralise. Also, the research suggests that fact and logic-based interventions have found to be efficient in many cases.

There is good news and bad news from the research on interventions. The good news is that, especially as a result of the misinformation campaigns threatening the integrity of the elections in 2016 (the Brexit campaign and the U.S. presidential election), the discussion of the possible counterstrategies is on a much more advanced level both in academic discussions and in policy discussions than before. The bad news is that experiences and academic research results are, in many cases, contradictory (e.g. on the usefulness of fact-checking), and it is difficult to put together a holistic picture from the very small and sporadic mosaics of research. A multiannual research project, based on meta-analyses of the existing research, replications and filling the existing gaps using a coherent theoretical framework and research design would be needed to have a more solid knowledge. As we live in an era of information warfare, there would be a need for a similar grand project to know more on propaganda and how to resist it than the one that was implemented during the biggest traditional war of the twentieth century: the Yale-program was focusing on the possibilities for attitude change and the resistance to persuasion in the context of the World Wars (for an overview, see McGuire 1996).

Note
1 Similarly, van Prooijen and van Dijk (2014) previously found that perspective taking can rather increase the receptivity to conspiracy theories.

References
Countering conspiracy theories


Countering conspiracy theories


