3
MISINFORMATION AND DISINFORMATION

Rachel Armitage and Cristian Vaccari

Introduction

This chapter will provide an overview of the debates surrounding misinformation and disinformation in political communication. After defining the key terms, we will contextualise this discussion, connecting recent concern about mis/disinformation with the growth of populism, declining trust in the media, and the role of digital platforms. We will then examine the key political, psychological, technological, and contextual factors that make people susceptible to mis/disinformation, before assessing current efforts to identify and tackle factually problematic content online. We conclude with broad recommendations for future efforts in this field.

Defining the key terms

Descriptors of problematic information range from general catch-all terms to specific references, competing in disputed hierarchies that change across people, context, and time. Terms including ‘junk news’ (Narayanan et al. 2018) and ‘fake news’ (Lazer et al. 2018) have variously been employed to represent manipulated, fabricated, extremist, satirical, sensationalist, parody, propagandist, and conspiratorial content (Tandoc, Lim & Ling 2018). This lack of definitional consistency potentially underlies conflicting academic findings (Tucker et al. 2018), illustrating how terminology can influence both the perception of the problem and its solution.

The most commonly accepted distinction between types of problematic information is misinformation and disinformation. Both terms refer to the sharing of incorrect, inaccurate, or misleading content, but they are separated by intentionality. While misinformation entails accidentally sharing inaccurate content, disinformation constitutes deliberate deception, often based on outright fabrications (Jack 2017). Difficult as it may be to determine the intentions of mis/disinformation sharers – especially when they are ordinary social media users – this distinction captures an important normative as well as empirical difference.

Some authors have suggested additional terms to cover conceptual space bordering mis/disinformation, including ‘xisinformation’, in which it is hard to parse a sharer’s intent (Jack 2017, 16), and ‘mal-information’ to describe intentionally harmful sharing of accurate information (Wardle & Derakhshan 2017, 5). However, this ongoing semantic discussion may distract from the need to tackle such problematic information (Weeks & Gil de Zúñiga 2019). Any such
efforts must start from a comprehensive understanding of the contextual factors that enable mis/disinformation to take root.

**Contextualising the challenge of mis/disinformation**

Whilst misleading information is not a new phenomenon, increasing concern about global mis/disinformation has revealed contemporary catalysts behind such problematic content (Lazer et al. 2018). One contributing factor is the rise of populism in political systems around the world. Populist leaders often attack experts and independent journalists, alleging bias in response to any negative media coverage (Newman et al. 2019). This is particularly effective in an increasingly polarised context, where populist supporters and detractors are ever less willing to engage with one another (Mason 2018). As a result, mis/disinformation has gained a stronger foothold in public discourse, reducing the space for evidence-based debate (Nyhan 2018). Political actors themselves often bear responsibility for spreading disinformation, especially during election campaigns. Prior to the United Kingdom’s 2016 referendum to leave or remain in the European Union (EU), there was a preponderance of misleading content favouring the Leave campaign on Twitter, with problematic assertions by the campaign itself achieving greater traction than disinformation efforts by outside groups (Gorrell et al. 2019). Similar activity was decried during the 2019 UK general election, with politicians accused of ‘playing fast and loose with the facts, avoiding journalistic scrutiny, and denigrating the media’ (Newman 2020, 12).

Accordingly, populism and political polarisation have contributed to the decline of trust in mainstream media and news organisations. In the UK, trust in news has been falling since 2015, and even the BBC is now seen as having an agenda, especially regarding divisive issues such as Brexit (Newman et al. 2019). Populist supporters are particularly likely to distrust independent news media, which they see as part of despised elites (Fawzi 2019), limiting the ability of the established media to authoritatively correct mis/disinformation. One possible outcome is the cultivation of an ‘anything-goes’ mentality among social media users (Chadwick & Vaccari 2019), who may become less vigilant about the quality of news they share as online social norms are eroded, and establishing the truth becomes increasingly difficult and contested.

Indeed, the broad proliferation of mis/disinformation has arguably been accelerated by social media (Wardle & Derakhshan 2017). Social networking sites have challenged the role of traditional media as information gatekeeper (Resnick, Ovadya & Gilchrist 2018), lowering the cost of entry to the production and distribution of news and thereby vastly increasing the quantity (but not necessarily the quality) of available content (Lazer et al. 2018). This has allowed politicians – especially those who can afford mass-scale digital advertising – to communicate directly with the public, devoid of the restrictions normally accompanying journalistic mediation (Siegel 2018). Social media has further facilitated the artificial inflation of problematic content via bots (automated accounts) and cyborgs (hybrid human/automated accounts), as well as empowering average citizens – including partisan activists – to quickly create and widely disseminate material of varying veracity (Cook, Ecker & Lewandowsky 2015), often using content from news organisations as a resource to influence others (Chadwick, Vaccari & O’Loughlin 2018). Mobile messaging apps (such as WhatsApp, Snapchat, Facebook Messenger, and Vine) pose distinctive challenges, facilitating private informal discussions between small, strong-tie networks that may be less guarded against mis/disinformation (Valeriani & Vaccari 2018). However, users are more likely to both issue and receive corrections when they share mis/disinformation on WhatsApp than they are on Facebook, possibly because they are less fearful of suffering backlash for challenging members of their strong-tie networks than they are when they engage with their weak ties on the more open social media platforms (Rossini et al. 2019).
Susceptibility to mis/disinformation

Exposure to mis/disinformation is widespread online. A survey of UK social media users found that as many as 57 percent believed they had seen inaccurate news on social media in the previous month, and another one-fifth of respondents could not tell whether they had or not (Chadwick & Vaccari 2019). During the COVID-19 pandemic, between half and one-quarter of the British public, depending on the period, reported coming across false or misleading news about the virus and an additional one-quarter was not sure (Ofcom 2020). Hence, it becomes important to identify factors that make individuals susceptible to mis/disinformation.

Political and social psychological factors

Humans respond to the uncertainty in our world through ‘shared sensemaking’, including sharing rumours and doubtful information (DiFonzo 2008, 10), which now occurs predominantly via social media and mobile messaging apps. Importantly, the framing of information under uncertainty can influence how we perceive a situation (Tversky & Kahneman 1986), as can our affinity for, or aversion to, uncertainty (Sorrentino & Roney 2000). Individuals favouring the familiar are likely to seek out certainty in societal groups (Lane 1962) that define their social identities (Hogg et al. 2007). Uncertainty reinforces feelings of in-group belonging and out-group enmity, creating inflated perceptions of differences between groups (Sherman et al. 2009). This is exacerbated if intergroup conflict is seen as a zero-sum game, in which gains for one group become losses for the other (Mason 2018).

Such intergroup conflict is arguably characteristic of party politics, particularly in majoritarian two-party systems in which there is little hope of cooperation across the aisle. To this end, exposure to messages reinforcing inter-party differences can stoke divisions and increase polarisation (Vaccari 2018), feeding into the creation of opposing shared realities (DiFonzo 2008). Indeed, as political groups are perceived as ever more disparate, they are simultaneously viewed by in-group members as ever more clearly defined (Sherman et al. 2009). In many political systems, personal identities and political affiliations have become increasingly aligned (Arceneaux & Vander Wielen 2017), creating socially homogenous parties with ever more intolerant members (Mason 2018), thus reinforcing feelings of ingroup belonging that validate one’s worldviews (Hogg et al. 2007).

Worldviews provide the basis for political ideologies, or the rationalised beliefs of a group employed by members in countering their opponents (Lane 1962). Individuals’ group identities therefore inevitably affect their political judgments, making it unlikely that they self-correct erroneous out-group beliefs for fear of undermining their worldview (Mason 2018). Instead people use heuristics, or stereotypes, concurrent with their worldview to cue their decisions (Lane 1962). Accordingly, messages alluding to group stereotypes can encourage acceptance of misleading information about political out-groups (Vaccari 2018; Nyhan 2018), discouraging critical reflection. Indeed, the term ‘political sectarianism’ has recently been proposed to more precisely capture the moralized nature of partisan identifications (Finkel et al. 2020). When political identities combine othering, aversion, and moral repulse towards other political groups, partisans become more willing to intentionally discount information that does not support their views and are even prepared to accept the use of anti-democratic tactics from their side if they can secure victory against opponents seen as immoral.

To this end, understanding the world mainly from a group perspective can contribute to social polarisation, or action based in prejudice and emotional volatility. People emotionally invest in the maintenance of their group identity even when this is irrational because
questioning an affiliation incorporating core aspects of one’s identity risks increasing uncertainty. In this vein, scholars contend that decision-making unavoidably begins with unconscious emotional intuition (Arceneaux & Vander Wielen 2017). Indeed, existing affect can prevent the rational processing of new information (Redlawsk 2002), and emotions are often the main catalyst for rumour transmission, overriding concerns for veracity (DiFonzo 2008; Lewandowsky et al. 2012). Individuals motivated to experience strong emotions need only to feel that they are accurate, forming strong opinions from gut reactions and failing to interrogate their intuition, and are thus particularly susceptible to believing and sharing mis/disinformation (Anspach, Jennings & Arceneaux 2019). While the mechanisms described here have existed for a long time, recent technological changes have arguably magnified their impact and implications.

**Technological factors**

The affordances of digital media exacerbate susceptibility to worldview-congruent mis/disinformation as they tend to facilitate instantaneous, uncritical behaviour based on little cognitive consideration (Bartlett 2018). Indeed, social media platforms design and constantly update their affordances to retain users’ attention, mainly by providing content they are more likely to interact with (Bucher 2018). Whilst users do have capacity to curate the content they see online, giving rise to concerns about echo chambers (Lewandowsky et al. 2012), research suggests that only a minority of users consistently avoid political content they disagree with (Vaccari & Valeriani forthcoming). Indeed, most social media users have ideologically diverse online networks and are therefore indirectly exposed to unfamiliar perspectives (Dubois & Blank 2018), although such exposure may reinforce rather than challenge users’ pre-existing beliefs (Bail et al. 2018) and has been associated with the sharing of mis/disinformation (Rossini et al. 2019).

The online environment is also attractive to nefarious actors running propaganda and disinformation campaigns (Narayanan et al. 2018), who take advantage of social media business models prioritising clicks and data capture over content quality. Such campaigns employ computational propaganda techniques, manipulating algorithms, cluttering conversations, and hijacking public spaces (such as hashtags) to reach users with inauthentic content (Sanovich & Stukal 2018). They further utilise microtargeting to tap into users’ identities, preferences, and prejudices, thus making it harder for users to recognise and reject mis/disinformation consistent with their beliefs (Bartlett 2018). Another concern is the rapid evolution of technologies that can distort audio-visual content, particularly in the creation of so-called deepfakes: synthetic videos generated to appear real by artificial intelligence software, increasingly available in open-source format and trained with publicly available data. Recent research suggests that even if individuals may not be misled by deepfakes, many react to them with uncertainty, reducing trust in all news encountered on social media as a result (Vaccari & Chadwick 2020).

**Contextual factors**

Political and media institutions can also affect the production, circulation, and impact of mis/disinformation. According to Humprecht, Esser & Van Aelst (2020), disinformation-resilient countries feature both an infrastructure that protects most citizens from exposure to false information (including strong public service broadcasters) and a citizenry less likely to believe and disseminate, and more likely to challenge, poor quality information. In an assessment of 18 Western democracies, the researchers found countries in Northern Europe (including Denmark, Finland, and The Netherlands) to be the most disinformation resilient, whilst Southern
European countries (including Italy, Spain, and Greece) and the United States were found to be particularly susceptible. This is confirmed by survey results showing greater concern about online mis/disinformation from citizens in South America (Brazil), the US, and the UK than from respondents in Germany and The Netherlands (Newman et al. 2019).

A nation’s resilience to mis/disinformation also depends on the type of social media platform favoured by its citizens. Notably, large WhatsApp and Facebook groups have become popular means of sharing and discussing news in non-Western countries such as Brazil, Malaysia, and Turkey. WhatsApp groups were exploited in both the 2018 Brazilian presidential election (Machado et al. 2019) and the 2019 Indian general election (Narayanan et al. 2019) to distribute misleading and divisive political content. Notably, information flows on mobile messaging apps are encrypted and generally private, making it hard for news organisations and digital platforms themselves to observe, correct, and limit the spread of inaccurate content via this medium.

Solutions to the problem of mis/disinformation

In light of the very real threat posed by mis/disinformation to democratic governance and the unprecedented scale at which digital media enable it to spread, efforts to effectively identify, limit, and correct it have become increasingly important.

Identifying mis/disinformation

To date, most work has focused on the development of automatic detection techniques to identify problematic content (Cook, Ecker & Lewandowsky 2015). These include classifiers to determine who might share misinformation online (Ghenai & Mejova 2018) and computational tools to detect clickbait, rumours, and bots. Whilst such tools can efficiently filter vast amounts of online content, the machine learning necessary to develop them can be prohibitively time and resource intensive (Tucker et al. 2018). Moreover, their effectiveness is limited by restricted access to social media data, as well as by increasingly effective automation that makes bots difficult to distinguish from real accounts. Even where detection is successful, bot developers soon adapt, and many bot accounts will never be discovered (Sanovich & Stukal 2018). Importantly, computational classifiers are not a replacement for human experts, who must intermittently retrain such tools and be on hand to mitigate any inadvertent discrimination classifiers may perpetuate (Ghenai & Mejova 2018). To this end, there is a risk that such tools will be more effective and precise in countries and on topics that digital platforms prioritise for investment.

Another approach in identifying mis/disinformation online is human content moderation. These efforts are conducted by large teams, employed by social media platforms to monitor and review content in line with their terms of service or community standards (Gillespie 2018). However, such work comes replete with associated mental health difficulties (Chen 2014), backlash from angry users (Newton 2019), and concerns about poor remuneration and the massive volume of content to be checked. Often moderators act based on user reporting, despite users’ tendency to rely on partisanship and ideology when identifying misleading content (boyd 2017) or, indeed, failing to challenge it all (Chadwick & Vaccari 2019). Finally, the often implicit assumptions underlying content moderation threaten to undermine the key democratic value of freedom of information (Kaye 2019). That any politically relevant but untrue speech should be censored and that such determinations should be made by quasi-monopolistic private companies arguably raises more problems than it solves.
Tackling mis/disinformation

Despite the difficulties inherent in isolating problematic content, further decisions must be taken about dealing with mis/disinformation when it is identified. Perhaps the easiest approach for platforms is to take down, downrank, or label such content, adjusting algorithms to promote reliable information, indicating source quality and story veracity, banning offending accounts, and removing harmful bot activity (Lazer et al. 2018). However, this kind of moderation risks bias or error such that legitimate content is accidentally removed, opening platforms up to censorship accusations (Sanovich & Stukal 2018). Moreover, previous attempts at algorithm change have counter-intuitively increased the prevalence of divisive topics (NewsWhip 2019), and there is no guarantee that algorithms cannot be gamed by bad actors. There is also the issue of what to do when disinformation is spread by popular political actors, including incumbent presidents, whom platforms have been more reluctant to police than ordinary users.

Another popular approach in tackling mis/disinformation is fact-checking. However, creating effective fact-checks is resource intensive, and their efficacy can be limited (Marwick 2018). The fractured nature of the online environment makes it difficult for corrections to reach those exposed to mis/disinformation, with problematic content going an average of 10 to 20 hours before fact-checking catches up (Shao et al. 2016). Where corrections do reach their desired audiences, repetition of mis/disinformation as part of the corrective effort merely increases its cognitive fluency – ‘the experience of ease or difficulty associated with completing a mental task’ (Oppenheimer 2008, 237) – paradoxically making acceptance of the misleading information more likely (Lazer et al. 2018). Indeed, if mis/disinformation is congruent with one’s worldview, belief in such content is especially likely to persist in the face of corrections (Cook, Ecker & Lewandowsky 2015). Moreover, problematic content often continues to influence attitudes even after corrections have been cognitively accepted (Thorson 2016). Nevertheless, various techniques can increase the effectiveness of correction, including an emphasis on facts (Lewandowsky et al. 2012), avoiding repetition of the mis/disinformation, issuing corrections as soon as possible, avoiding negation, reducing ideological or partisan cues, citing credible sources, using graphics (Nyhan & Reifler 2012), and providing causal alternatives (Thorson 2016). However, even such carefully crafted corrections have a greater chance of success if they are congruent with their target’s worldview (Lewandowsky et al. 2012). Accordingly, misleading political content is particularly difficult to correct (Walter & Murphy 2018), with journalists finding that their fact-checks are failing to impact large portions of the public (Newman 2020).

An alternative tool in the fight against mis/disinformation is legislation. In Britain, recent high-level reports have called for the independent regulation of social media companies, with legal and financial consequences for failing to protect users against harmful content (Digital, Culture, Media & Sport Committee 2019; Cairncross et al. 2019), although other expert groups have recommended against such measures (Council of Europe 2018; European Commission 2018; Nielsen et al. 2019). There is a fine line between controlling problematic information and compromising free speech, with any government regulation efforts potentially facing criticism for censorship or partiality. The lack of an agreed definition for problematic content (boyd 2017) and, perhaps more importantly, of a shared understanding of the problem across the political spectrum also constitute major obstacles to regulatory solutions.

Other efforts to tackle mis/disinformation have focused on changing journalistic practice. A key concern here is the decline of resources among news organisations, which has led many outlets, especially at the local level, to close or substantially downscale newsrooms. However, many journalists feel their industry should make greater efforts to challenge misleading information (Newman 2020) and attempt to re-establish their reputations – possibly with some help
from civil society organisations that are developing training programmes on source verification and responsible reporting (First Draft n.d.). However, efforts to restore public trust in the media may demand more financial and editorial resources than are available (Jukes 2018), and continued efforts to fact check will be perpetually undermined or ignored by partisan actors – without whose self-restraint journalists will always be facing an uphill battle.

Approaches focused on news consumers are also relevant, with civic and digital education efforts seeking to equip social media users against online mis/disinformation. Research suggests that young people are often unable to tell real from misleading news content on social media (McGrew et al. 2017) and that media literacy strategies – for instance, teaching users how to employ fact-checking techniques – could help address this (Wineburg & McGrew 2017). Attempts to inoculate users against mis/disinformation have also shown promise, as ‘fake news’ games have often improved identification of, and resistance to, misleading content (Roozenbeek & van der Linden 2019). Nevertheless, there are concerns that media literacy techniques may be ineffectual (Marwick 2018), with questions about the long-term maintenance of skills and the risk of such efforts reducing trust in news altogether (Lazer et al. 2018). Media literacy is also time and resource intensive and is unlikely to reach or affect all users who need it.

Finally, changes in the way social media works may also be helpful. For instance, creating more friction in the user experience may reduce users’ inclination to mindlessly accept and share mis/disinformation that fits with their worldview. Research suggests that inducing social media users to think about accuracy, or to think critically, can reduce the likelihood that users trust, like, or share problematic information online (Pennycook et al. 2019). Requiring defence of a decision can trigger accuracy motivations (Kunda 1990), and this in turn encourages successful processing of worldview-incongruent information (Redlawsk 2002). Moreover, simply introducing elements of difficulty or ‘disfluency’ (unease) to a task can interrupt and correct reasoning based on partisan cues (Alter et al. 2007). Indeed, interventions based on disfluency have been found to increase analytic thinking and to reduce belief in conspiracy theories (Swami et al. 2014). If such friction can be introduced into the social media environment, it might reduce users’ propensity to believe and share mis/disinformation, although the question of how to achieve this in practice remains to be answered. This is a momentous task, as the success of social media platforms has largely depended on encouraging fluency, ease of use, and ‘stickiness’ – the ‘ability to attract users, to get them to stay longer, and to make them return again and again’ (Hindman 2018, 2).

Conclusion

This chapter has sought to define, contextualise, and explain susceptibility to online mis/disinformation and to discuss solutions to the problem. These phenomena are rooted in psychological, political, technological, and contextual factors that are unlikely to change substantially in the future, and there is no single effective approach that might limit or reverse this course. Thwarting the capability of disinformation campaigns – whether run by mainstream political actors, outside groups, or foreign governments – to exploit a continually changing online landscape and to employ social media users as their (knowing or unknowing) foot soldiers in spreading problematic content will be no easy task.

Whilst there is no apparent ‘silver bullet’ solution to mis/disinformation, such content relies on exposure to and engagement by average social media users to be effective. Therefore, whilst it is undoubtedly important to continue developing tools and techniques that take on the production and distribution of misleading information, there is surely a key, long-term role to be
played by approaches seeking to equip citizens with the tools to recognise, challenge, and reject mis/disinformation. However, even this might prove insufficient if the underlying political and technological catalysts for such content cannot be remedied.

Bibliography


Rachel Armitage and Cristian Vaccari


46


