

Misinformation and trusted voices

Bakir, Vian; McStay, Andrew

Published: 12/10/2022

Publisher's PDF, also known as Version of record

[Cyswllt i'r cyhoeddiad / Link to publication](#)

Dyfyniad o'r fersiwn a gyhoeddwyd / Citation for published version (APA):

Bakir, V., & McStay, A. (2022, Oct 12). Misinformation and trusted voices: Addressing false information online via provision of authoritative information: Why dialling down emotion is part of the answer. UK Parliament.

Hawliau Cyffredinol / General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Written evidence submitted by Professor Vian Bakir and Professor Andrew McStay

Addressing false information online via provision of authoritative information: Why dialling down emotion is part of the answer

Invited Submission by Vian Bakir & Andrew McStay to DCMS Online Harms and Disinformation Sub-Committee Inquiry into Misinformation and Trusted Voices

September 2022

Vian Bakir is Professor of Journalism & Political Communication, Bangor University, Wales, UK.
Andrew McStay is Professor of Digital Life, Bangor University, Wales, UK.

Both are from the [Emotional AI Lab](#) and the [Network for Study of Media & Persuasive Communication](#). Both are authors of [Optimising Emotions, Incubating Falsehoods: How to Protect the Global Civic Body from Disinformation and Misinformation](#). Springer (2022).

Section 1. Summary

1.1 We primarily address the Inquiry's q.6: 'Is the provision of authoritative information responsive enough to meet the challenge of misinformation that is spread on social media?'

1.2 Drawing on multi-disciplinary scholarship, we advise that any solution considered for countering the spread of false information online (including the solution of providing authoritative information) should be mindful of the many types of actors and communicative processes in play. In Section 2 we outline the complexities generated by diverse actors and numerous communicative processes (namely philosophical, epistemological, cultural, political, economic, psychological, social and technological). In Section 3 we consider what these insights imply for the solution of providing authoritative information to counter the spread of false information online.

1.3 In Section 4, we conclude from our analysis of diverse actors and of pertinent philosophical, epistemological, cultural, political, economic, and psychological communicative processes, that the solution of providing authoritative information to address the spread of false information online:

- Will not sway people who already do not trust that information (content, source, or channel), or who spread false information in order to express their group identity or dissatisfaction with the political system;
- Could prove useful for the undecided or confused if presented in an understandable fashion through trusted routes. Such provision would be most effective on issues where people have not yet made up their mind rather than presented as a corrective to false information. Such provision can be encouraged through better financing of in-depth, impartial journalism; and provision of plain English overviews of the pattern of expert (e.g. scientific) consensus on any issue.

1.4 Additionally, and drawing from our analysis of social and technological communicative processes, we conclude that rather than having to make difficult content moderation decisions about what is true and false on the fly and at scale, it may be better to ensure that digital platforms' algorithms *optimise emotions for social good* rather than just for the platform and its advertisers' profit. What this social good optimisation would look like is worthy of further study, but we posit that this would likely involve *dialling down the platform's emotional contagion, and engagement, of users*.

Section 2. A complex array of actors and communicative process

2.1 Any solution considered for countering the spread of false information online (including the solution of providing authoritative information) should be mindful of the many types of actors and communicative

processes in play in spreading misinformation (i.e. inadvertently inaccurate information) or disinformation (i.e. deliberately inaccurate information).

2.2 On **actors**, audiences are diverse, and what is perceived as ‘authoritative information’ depends on factors such as the specific audience member’s political leanings,¹ and trust in the communicators,² in the media outlet,³ and in ‘the system’.

For instance, experiments show that US participants with higher levels of populist attitudes, media distrust, and fake news perceptions are more likely to find established information untrustworthy and more likely to find misinformation credible.⁴ Recent European-based comparative research across 10 countries also finds that people with stronger populist attitudes tend to believe that most news media spread misinformation and, especially, disinformation.⁵ A survey-based study of the characteristics of the audiences of right-wing alternative online media across Northern and Central Europe finds that such audiences are sceptical of news quality in general, particularly distrust public service broadcasting media, and use social media as a primary news source.⁶

Not just the preserve of populists and the alt-right, international surveys drawn from every continent find that in 2022, only 42% of people trust the news most of the time: in the UK this figure is lower, at just 34% (a downward trend, as the figure was 51% in 2015). Even the most trusted news outlets in the UK (namely, public broadcasters that are required to meet strict impartiality standards), garner trust from only just over 50% of the British population.⁷

2.3 On **communicative processes**, there are many that complicate the provision of authoritative information as a solution to false information online. These communicative processes are as follows:

2.3.1 *Philosophical and epistemological.* The rise of relativism, and claims of ‘alternative facts’ and ‘fake news’ can make it hard to agree what constitutes ‘being authoritative’, with some suggesting that we now live in a ‘post-truth’ world where appeals to opinions and emotion matter more than facts, or where the status of facts is downgraded to that of mere opinion.⁸ Whether people perceive false information to be inadvertently inaccurate (i.e. misinformation) or deliberately inaccurate (i.e. disinformation) is also pertinent: for instance a 10-country EU-based study finds that people with disinformation perceptions do not view political institutions as responsible or capable of dealing with said disinformation; while those with misinformation perceptions are more supportive of political interventions to check the veracity of online information.⁹

2.3.2 *Cultural.* There has been a decline of trust in key institutions evidenced in many surveys from the first decade of the 21st century in the UK, EU, USA and Australia. UK-based surveys identify industry officials, government officials and journalists as ranking among the lowest on the trust scale, and this pattern

¹ Hameleers, M., Brosius, A., Marquart, F., Goldberg, A. C., van Elsas, E., & de Vreese, C. H. (2021). Mistake or manipulation? Conceptualizing perceived mis- and disinformation among news consumers in 10 European countries. *Communication Research*. <https://doi.org/10.1177/0093650221997719>

² Sterrett, D., Malato, D., Benz, J., Kantor, L., Tompson, T., Rosenstiel, T., Sonderman, J., & Loker, K. (2019). Who shared It? Deciding what news to trust on social media, *Digital Journalism*, 7(6), 783-801. <https://doi.org/10.1080/21670811.2019.1623702>

³ Shehata, A., & Strömbäck, J. (2022). Media Use and Societal Perceptions: The Dual Role of Media Trust. *Media and Communication*, 10(3), 146-157. doi: <https://doi.org/10.17645/mac.v10i3.5449>

⁴ Hameleers, M (2022) “I don’t believe anything they say anymore!” Explaining unanticipated media effects among distrusting citizens. *Media & Communication*, 10(3). <https://doi.org/10.17645/mac.v10i3.5307>

⁵ Hameleers, M., Brosius, A., Marquart, F., Goldberg, A. C., van Elsas, E., & de Vreese, C. H. (2021). Mistake or manipulation? Conceptualizing perceived mis- and disinformation among news consumers in 10 European countries. *Communication Research*. <https://doi.org/10.1177/0093650221997719>

⁶ Schulze, H. (2020). Who uses right-wing alternative online media? An exploration of audience characteristics. *Politics and Governance*, 8(3), 6–18. <https://doi.org/10.17645/pag.v8i3.2925>

⁷ Reuters Institute digital news report (2022). <https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2022/united-kingdom>

⁸ Farkas, J., & Schou, J. (2020). *Post-truth, Fake News and Democracy: Mapping the Politics of Falsehood*. Routledge. Van Aelst, P., Stromback, J., Aalberg, T., Esser, F., de Vreese, C.H., Matthes, J., Hopmann, D., Salgado, S., Hub., N., Stepińska, A., Papathanassopoulos, S., Berganza, R., Legnante, G., Reinemann, C., Sheaffer, T., & Stanyer, J. (2017). Political communication in a high choice media environment: A challenge for democracy? *Annals of the International Communication Association*, 4, 3–27. <https://doi.org/10.1080/23808985.2017.1288551>

⁹ Hameleers, M., Brosius, A., Marquart, F., Goldberg, A.C., van Elsas, E., & de Vreese, C.H. (2021). Mistake or manipulation? Conceptualizing perceived mis- and disinformation among news consumers in 10 European countries. *Communication Research*. <https://doi.org/10.1177/0093650221997719>

continues today. In general, deference to authority has also declined and people are less willing to unquestioningly accept government or expert advice. Two decades ago, reasons proffered for this in the UK included the power of local knowledge; the rise in individualism; distrust in the wake of publicised past mistakes on public safety issues; and corruption and conflicts of interest among political elites.¹⁰

Conversely, consistently among the most trusted professions in the UK are scientists. The UK public is generally positive towards science and scientists, with over 80% trusting them to tell the truth in MORI's polls across 2019-2021.¹¹ A more detailed poll from 2020 finds that 60% consider scientists in general to be trustworthy, but those from social class C2DE (the less affluent) and non-graduates tend to be less positive and less trusting than the middle classes and graduates. Also less trusting are those who are sceptical about the benefits of science.¹² Reasons for why people may distrust scientists include a privileging of individual, everyday lay experience and common sense (or even gut feeling) over scientific method (itself often replete with uncertainties, disagreements, data limitations and multiple interpretations of evidence), as well as a perceived lack of evidence.¹³ Indeed, in 2019, 18% of UK adults had low scores on 'science capital' (such as science-related qualifications, knowledge, contacts, informal science learning and scientific literacy), this being far higher among those with no qualifications (61%). Furthermore, 65% of UK adults believe that there is so much conflicting information about science that it is difficult to know what to believe.¹⁴

2.3.3 *Political*. Ruling cultures of spin, deception, bullshit (an accepted academic term) and corruption are likely to have corrosive effects on people's trust in government.¹⁵ Two decades ago, reasons proffered for a decline in trust in government in the UK included government misinformation and pro-active government news management strategies. For instance, government spin generated adversarial media responses, leading the public to expect the worst of politicians, even when evidence supports the government's position.¹⁶ Two decades later, today in the UK, only 20% think that the media are independent from undue political or government influence or from undue business or commercial influence.¹⁷ More globally, a survey in 2020 across 40 countries finds that it is domestic politicians that are seen as by far the most responsible for false and misleading information online (40%), followed by political activists (14%), journalists (13%), ordinary people (13%), and foreign governments (10%).¹⁸ Furthermore, studies suggests that sharing fake news might be an expression of group identity or dissatisfaction with the current political system.¹⁹

¹⁰ Bakir, V. & Barlow, D. (2007). The age of suspicion. In Bakir, V. & Barlow, D. (eds). *Communication in the Age of Suspicion: Trust and the Media*. Palgrave Macmillan.

¹¹ Ipsos MORI (2019). *Ipsos MORI Veracity Index 2019*. <https://www.ipsos.com/en-uk/trust-politicians-falls-sending-them-spiralling-back-bottom-ipsos-mori-veracity-index>. Ipsos MORI (2020). *Ipsos MORI Veracity Index 2020*. <https://www.ipsos.com/en-uk/ipsos-mori-veracity-index-2020-trust-in-professions>. Ipsos MORI (2021). *Ipsos MORI Veracity Index 2021*.

https://www.ipsos.com/sites/default/files/ct/news/documents/2021-12/trust-in-professions-veracity-index-2021-ipsos-mori_0.pdf

¹² Skinner, G., Garrett, C. & Shah, J.N. (2020). *How has COVID-19 affected trust in scientists?* <https://www.ukri.org/wp-content/uploads/2020/09/UKRI-271020-COVID-19-Trust-Tracker.pdf>

¹³ Mede, N.G. & Schäfer, M.S. (2020). Science-related populism: Conceptualizing populist demands toward science. *Public Understanding of Science*, 29(5), 473–491. <https://doi.org/10.1177/0963662520924>

259. Sleight, C. (2021). *Fluoridation of drinking water in the UK, c.1962-67. A case study in scientific misinformation before social media*. The Royal Society, <https://royalsociety.org/-/media/policy/projects/online-information-environment/oie-water-fluoridation-misinformation.pdf>. Department for Business, Energy and Industrial Strategy (2020). *Public attitudes to science 2019*, Main report BEIS Research Paper Number 2020/012,

https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/905466/public-attitudes-to-science-2019.pdf

¹⁴ Department for Business, Energy and Industrial Strategy (2020). *Public attitudes to science 2019*, Main report BEIS Research Paper Number 2020/012, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/905466/public-attitudes-to-science-2019.pdf

¹⁵ Bakir, V., Herring, E., Miller, D. & Robinson, P. (2018). Lying and Deception in Politics. In J. Meibauer (Ed.), *The Oxford Handbook of Politics and Lying* (pp. 529-540). Oxford University Press. Bakir, V., Herring, E., Miller, D. & Robinson, P. (2018). Organized persuasive communication: A new conceptual framework for research on public relations, propaganda and promotional culture. *Critical Sociology*, 45(3), 311-328. <https://doi.org/10.1177/0896920518764586>

¹⁶ Bakir, V. & Barlow, D. (2007). The age of suspicion. In Bakir, V. & Barlow, D. (eds). *Communication in the Age of Suspicion: Trust and the Media*. Palgrave Macmillan.

¹⁷ Reuters Institute digital news report (2022). <https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2022/united-kingdom>

¹⁸ Reuters Institute digital news report (2020).

https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2020-06/DNR_2020_FINAL.pdf

¹⁹ Nisbet, E.C. & Kamenchuk, O. (2019). The psychology of state-sponsored disinformation campaigns and implications for public diplomacy. *The Hague Journal of Diplomacy*, 14, 65-82. <https://doi.org/10.1163/1871191X-11411019>

2.3.4 *Economic*. There are increasing competitive pressures on news outlets as their income has declined in the digital age. Fewer people are willing to pay for news (only 9% of UK users pay for online news in 2022²⁰), and digital advertisers eschew news outlets in favour of the Facebook-Google duopoly. This damages the news product, for instance, forcing regional and local news outlets to close or restructure, and pushing remaining news outlets to be more reliant on press releases rather than engaging in more expensive on the ground or investigative journalism.²¹ Economic pressures have also generated a tendency among digital news outlets to produce content designed to attract online users and generate revenue, and as such, appealing to dominant digital platforms' algorithms became vital for the economic survival of news outlets: these algorithms prioritise content that is engaging rather than factual or truthful.²² Also note that fact-checking by independent fact-checkers is resource-intensive, and hence expensive and used very selectively.²³

2.3.5 *Psychological*. A meta-analysis of the psychological efficacy of messages countering misinformation finds that debunking effects are weaker when audiences generate reasons in support of the initial misinformation, indicating the operation of 'confirmation bias' (where people unwittingly seek or interpret information in ways that conform with their existing beliefs or hypotheses).²⁴ Correcting misinformation therefore does not necessarily change people's attitudes and beliefs.²⁵ Experiments also show that repeated exposure to fake news headlines increases their perceived accuracy: this occurs despite a low level of overall believability, and even when stories are labelled as contested by fact-checkers or are inconsistent with readers' political ideology. These results suggest that platforms help incubate belief in false information (by allowing false information to go viral), and that tagging such stories as 'disputed' is ineffective as any repetition of misinformation, even in the context of refuting it, may be harmful.²⁶ Even independent fact-checking (the practice of systematically publishing assessments of the validity of claims made by public bodies to identify whether a claim is factual) are not necessarily influential. A meta-analysis of 30 studies finds that although people's beliefs become more accurate and factually consistent after exposure to a fact-checking message, the fact-checking has weak impacts on beliefs that become negligible the more the study resembles real-world scenarios of exposure to fact-checking.²⁷

However, psychological research also shows that inoculating people with information *before* their minds are made up on an issue may better ensure that false information does not circulate. Recent studies find that inoculating people with facts against misinformation works for a highly politicised issue (global warming), regardless of prior attitudes.²⁸ Applying inoculation theory to fake news finds that inoculation has some effect in making participants more sceptical, and attuning people to deception.²⁹

2.3.6 *Social and technological*. This refers to how interaction between society and technology creates unique points of interest. We address four pertinent features, below, of the interactions between technology and people evident in the contemporary media 'ecology'.

²⁰ Reuters Institute digital news report (2022). <https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2022/united-kingdom>

²¹ Nielsen, R.K. & Fletcher, R. (2020). Democratic creative destruction? The effect of a changing media landscape on democracy. In N. Persily & J.A. Tucker (Eds.), *Social Media and Democracy: The State of the Field and Prospects for Reform*, (pp. 139-162). Cambridge University Press

²² Bakir, V. & McStay, A. (2022). *Optimising Emotions, Incubating Falsehoods: How to Protect the Global Civic Body from Disinformation and Misinformation*. Springer.

²³ Oshikawa, R., Qian, J., & Wang, W.Y. (2020). A survey on natural language processing for fake news detection. *Proceedings of the 12th Language Resources and Evaluation Conference (LREC 2020)* pp. 6086-6093. <https://arxiv.org/pdf/1811.00770.pdf>

²⁴ Chan, M.S., Jones, C.R., Jamieson, K.H., & Albarracín, D. (2017). Debunking: A meta-analysis of the psychological efficacy of messages countering misinformation. *Psychological Science*, 28(11), 1531-1546. doi: 10.1177/0956797617714579

²⁵ Flynn, D.J., Nyhan, B. & Reifler, J. (2017). The nature and origins of misperceptions: Understanding false and unsupported beliefs about politics. *Political Psychology*, 38(5), 127-150. <https://doi.org/10.1111/pops.12394> . Wittenberg, C. & Berinsky, A.J. (2020).

Misinformation and its correction. In N. Persily & J.A. Tucker (Eds.), *Social Media and Democracy: The State of the Field and Prospects for Reform*, (pp. 163-198). Cambridge University Press.

²⁶ Pennycook, G., Cannon, T.D., & Rand, D. G. (2018). Prior exposure increases perceived accuracy of fake news. *Journal of Experimental Psychology: General*, 147(12), 1865-1880. <https://doi.org/10.1037/xge0000465>

²⁷ Walter, N., Cohen, J., Holbert, R.L. & Morag, Y. (2020). Fact-checking: A meta-analysis of what works and for whom, *Political Communication*, 37(3), 350-375. <https://doi.org/10.1080/10584609.2019.1668894>

²⁸ Cook, J., Lewandowsky, S. & Ecker, U.K.H. (2017). Neutralizing misinformation through inoculation: Exposing misleading argumentation techniques reduces their influence. *PLOS ONE*, 12(5), 1–21. <https://doi.org/10.1371/journal.pone.0175799>. van der Linden, S., Leiserowitz, A., Rosenthal, S. & Maibach, E. (2017). Inoculating the public against misinformation about climate change. *Global Challenges*, 1(2). <https://doi.org/10.1002/gch2.201600008>

²⁹ Roozenbeek, J. & van der Linden, S. (2019). Fake news game confers psychological resistance against online misinformation, *Palgrave Communications*, 5, (65). <https://doi.org/10.1057/s41599-019-0279-9>

Ready exposure to non-factual and opinionated information. Our ‘high choice’ digital media ecology allows ready exposure to alternative media, partisan media, ideological media, and influencers, none of which are primarily guided by journalistic norms or standards, but rather operate to counter information in mainstream media, confirm the worldviews and attitudes of their targeted audiences, and attract followers.³⁰

People are fooled by new deceptive forms. For instance, studies on users’ responses to the rise of AI-generated deepfakes find limited capacity to recognise this new deceptive form, especially when the content presented is neutral rather than suspiciously out of character.³¹

False information is viral on social media. Big data studies (on Twitter) demonstrate that false information is contagious online, with falsehood diffusing significantly farther, faster, deeper, and more broadly than the truth;³² with fact-checking content typically lagging that of misinformation or false rumours by 10-20 hours;³³ and with low credibility content equally or more likely to spread virally as fact-checked articles.³⁴

Emotions are viral on social media. Multiple big data studies find that expression of emotion is socially contagious on social media (meaning that a perceiver’s emotions become more similar to others’ emotions as a result of exposure to these emotions), with caveats that causality is difficult to prove.³⁵ For instance, a computational, comparative study of Italian Facebook pages’ reporting on two polarised communities (scientific and conspiracy) across 2010-2012 shows that in both communities, more negative emotional states (ascertained by sentiment analysis of users’ posts) is driven by more frequent posting of comments.³⁶ Another study finds that presence of moral-emotional words on Twitter in three polarising issues increased their transmission by 20% per word.³⁷ Such emotional contagion is not an accident, but the result of digital platforms’ business model where algorithms are constantly tweaked to maximise user engagement.³⁸

Section 3. Implications for the solution of providing authoritative information to address false information online

3.1 People’s views, judgements and disagreements flourish on social media. This is a good thing for civic engagement and democratic debate, especially on issues of national importance where the knowledge base is uncertain. However, false content proliferates from a confluence of the practices of disinformation actors (including domestic politicians), what people find engaging (often emotional or false content), and the algorithms of dominant platforms (designed to encourage user engagement for financial gain). Countering this with ‘authoritative information’ is not straightforward as audiences are diverse, and what is perceived as

³⁰ Benkler, Y. (2020). A political economy of the origins of asymmetric propaganda in American media. In W.L Bennett & S. Livingston (Eds.), *The Disinformation Age: Politics, Technology and Disruptive Communication in the Information Age* (pp. 43-66). Cambridge University Press. doi.org/10.1017/9781108914628. Figenschou, T.U. & Ihlebaek, K.A. (2019). Challenging journalistic authority: Media criticism in far-right alternative media. *Journalism Studies*, 20(9), 1221–1237. Shehata, A. & Strömbäck, J. (2022). Media Use and societal perceptions: The dual role of media trust. *Media and Communication*, 10(3), 146-157. doi:https://doi.org/10.17645/mac.v10i3.5449

³¹ Vaccari, C. & Chadwick, A. (2020). Deepfakes and disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and trust in news. *Social Media + Society*, 6(1). https://doi.org/10.1177/2056305120903408. Lewis, A., Vu, P., Duch, R. M. & Chowdhury, A. (2022). *Do content warnings help people spot a deepfake? Evidence from two experiments*. The Royal Society, https://royalsociety.org/-/media/policy/projects/online-information-environment/do-content-warnings-help-people-spot-a-deepfake.pdf

³² Vosoughi, S., Roy, D. & Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380), 1146-1151. DOI: 10.1126/science.aap9559

³³ Shao, C., Ciampaglia, G. L., Flammini, A. & Menczer, F. (2016). Hoaxy: A platform for tracking online misinformation. *Proceedings of the 25th international conference companion on world wide web* (pp. 745-750). https://arxiv.org/abs/1603.01511.

Zubiaga, A., Liakata, M., Procter, R., Wong Sak Hoi, G., & Tolmie, P. (2016) Analysing how people orient to and spread rumours in social media by looking at conversational threads. *PLoS ONE*, 11(3), e0150989. doi:10.1371/journal.pone.0150989

³⁴ Shao, C., Ciampaglia, G.L., Varol, O., Yang, K.C., Flammini, A., & Menczer, F. (2018). The spread of low-credibility content by social bots. *Nature Communications*, 9, 4787. https://doi.org/10.1038/s41467-018-06930-7

³⁵ Goldenberg, A. & Gross, J. J. (2020). Digital emotion contagion. *Trends in Cognitive Sciences*, 24(2), 316-

328. https://doi.org/10.1016/j.tics.2020.01.009 . Kramer, A.D.I., Guillory, J.E., & Hancock, J.T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences*, 111 (29), 8788–90. https://doi.org/10.1073/pnas.1320040111.

³⁶ Del Vicario, M., Vivaldo, G., Bessi, A., Zollo, F., Scala, A., Caldarelli, G. & Quattrociocchi, W. (2016). Echo chambers: emotional contagion and group polarization on Facebook. *Nature*, 6, 37825. DOI: 10.1038/srep37825.

³⁷ Brady, W.J., Wills, J.A., Jost, J.T., Tucker, J.A., Van Bavel, J.J. (2017). Moral contagion in social networks. *Proceedings of the National Academy of Sciences*, 114 (28), 7313-7318. https://doi.org/10.1073/pnas.1618923114.

³⁸ Bakir, V. & McStay, A. (2022). *Optimising Emotions, Incubating Falsehoods: How to Protect the Global Civic Body from Disinformation and Misinformation*. Springer.

‘authoritative information’ depends on factors such as the specific audience member’s political leanings, and trust in the communicators, the media outlet and ‘the system’. With trust in mainstream media currently quite low in the UK, and very low in Europe among certain types of people (e.g. populists and the alt-right), great care would be needed to avoid unintended consequences. The provision of authoritative information will also need to sit alongside other views (as censorship goes against democratic norms).

3.2 Considering *philosophical, epistemological, cultural, political* and *psychological* communicative processes, the provision of authoritative information will not sway people who already do not trust that information (content, source, or channel), or who are sharing false information as an expression of group identity or dissatisfaction with the political system. At worst, such provision of authoritative information will simply be regarded as another manipulative ploy by untrustworthy elites (especially politicians) and experts to bolster their worldview and suppress alternatives.

However, for people who do not hold strong views or who have not made up their mind on an issue, or who are confused by conflicting information, provision of authoritative information could prove useful if presented in an understandable fashion through trusted routes. The provision of authoritative information would be most effective on issues where people have not yet made up their mind rather than presented as a corrective to false information. However, if seeking to correct false information, it would also need to take care not to repeat the false information that it is seeking to counter.

3.3 Considering *economic* communicative processes, that authoritative facts are expensive to resource (be this via newspapers engaging in investigative journalism or fact-checkers checking viral content) means that the supply and circulation of authoritative facts will be vastly outstripped by the supply of opinion, misinformation and disinformation, especially as non-factual material can be as emotive, engaging and false as it likes. To help address this, and also considering what news outlets the majority of people trust, the provision of authoritative information would need to be encouraged by financially supporting news outlets to engage in in-depth, impartial journalism.

Another conduit of authoritative facts is the experts themselves (e.g. academics). In the UK, they are already encouraged (and hence financed) by their university employers and research funding councils to communicate their research to the public and the press. However, presenting their findings, the methodology, the limitations of the study, and how this benefits society in plain English can be challenging, let alone in ways that might be engaging. Furthermore, academic knowledge generally progresses incrementally, whether via theoretical critique or empirical hypothesis testing, with much disagreement about appropriateness of methods and interpretation of significance of results. Airing such slow-motion disagreement in public could be misinterpreted by those lacking in scientific literacy, or who do not have ready access to plain English syntheses that can provide an overview of the pattern of expert consensus. Also, although the scholarly peer review process would normally weed out weakly supported, cherry-picked findings, this would not prevent weak, unpublished findings from being aired on social media.

3.4 Rather than having to make difficult content moderation decisions (be this by platforms and their AIs, or by digital regulators) about what is true and false on the fly and at scale, it may be better to ensure that algorithms *optimise emotions for social good* rather than just for the platform and its advertisers’ profit. What this social good optimisation would look like is worthy of study in itself,³⁹ but we posit that this would likely involve *dialling down the platform’s emotional contagion, and engagement, of users*. However, this goes against the dominant platforms’ business model, and most governments express reluctance to stifle innovation in the technology industry.⁴⁰ Without action in these areas, we are largely left with the solution of improving people’s ability to recognise false information online. This would involve improving a range of literacies including digital literacy (e.g. awareness of the forms of falsehood online, and of algorithmic prioritisation of emotional content) as well as people’s scientific capital (mentioned earlier). This remains a difficult area, however, for all the reasons we have discussed in this submission.

Section 4. Conclusion

³⁹ This is something that we at the Emotional AI Lab are turning our minds to. In particular, see McStay, A. (2022). *Automating empathy: When technologies claim to feel-into everyday life*. Oxford University Press.

⁴⁰ Bakir, V. & McStay, A. (2022). *Optimising Emotions, Incubating Falsehoods: How to Protect the Global Civic Body from Disinformation and Misinformation*. Springer.

4.1 Drawing from our analysis of diverse actors and of pertinent *philosophical, epistemological, cultural, political, economic* and *psychological* communicative processes, we conclude that the solution of providing authoritative information to address the spread of false information online will not sway people who already do not trust that information (content, source, or channel), or who spread false information in order to express their group identity or dissatisfaction with the political system. However, for people who do not hold strong views or allegiances, or who have not made up their mind on an issue, or who are confused by conflicting information, improving the supply of authoritative information online could prove useful if presented in an understandable fashion through trusted routes. Such provision would be most effective on issues where people have not yet made up their mind rather than presented as a corrective to false information. Such provision can be encouraged through better financing of in-depth, impartial journalism; and provision of plain English overviews of the pattern of expert (e.g. scientific) consensus on any issue.

4.2 Additionally, drawing from our analysis of *social* and *technological* communicative processes, rather than having to make difficult content moderation decisions about what is true and false on the fly and at scale, it may be better to ensure that digital platforms' algorithms *optimise emotions for social good* rather than just for the platform and its advertisers' profit. What this social good optimisation would look like is worthy of further study, but we posit that this would likely involve *dialling down the platform's emotional contagion, and engagement, of users*. Without action on this core socio-technological area, we are largely left with the solution of improving people's ability to recognise false information online. This would involve improving a range of literacies including digital literacy (e.g. awareness of the forms of falsehood online, and of algorithmic prioritisation of emotional content) and scientific literacy. This remains a difficult area, however, for all the reasons we have discussed in this submission.