Date received: 20 May 2024

Email: A.Trofimov-692@kent.ac.uk

# Applying Evidential Pluralism to Justify Legal Responses to Online Fake News

Alexandra Trofimov

Abstract:

The aim of this paper is to argue that Evidential Pluralism ought to be used to evaluate the impact of online fake news. To support this, I show how an application of Evidential Pluralism can overcome difficulties in assessing the impact of online fake news. The significance of this is twofold. Firstly, the application of Evidential Pluralism enables an evidence-based justification for legal interventions aimed at tackling online fake news. Secondly, the application of Evidential Pluralism to the problem of online fake news provides a case study example to motivate a new methodology for evidence-based law, called EBL+.

Key Words:

Fake News; The Justification of Laws; Evidence-Based Law; Evidential Pluralism; EBL+

## 1. Introduction

The ease with which false and misleading information can be disseminated online has led to widespread concerns about the detrimental effects of online fake news, including concerns

about democratic stability and public health security. These concerns have led numerous jurisdictions to propose or implement legal interventions to mitigate the impact of online fake news.

Until recently, however, little evidence has been provided to support the concern that online fake news has a real and negative impact. This lack of evidence is problematic for the justification of legislative and regulatory interventions. If online fake news has minimal or no negative impact, then the necessity and proportionality of legislative and regulatory interventions is undermined.

The lack of evidence to support the concern that online fake news has a negative impact is due, in no small part, to the difficulty in isolating and assessing the extent of the impact of online fake news. These difficulties, however, can be overcome.

The aim of this paper is to argue that Evidential Pluralism ought to be used to evaluate the impact of online fake news. Evidential Pluralism is a philosophical account of causal enquiry according to which establishing a causal claim requires establishing both correlation and mechanisms. As we shall see, combining evidence of correlation and evidence of mechanisms enables robust overall evidence of the effects of online fake news to be provided. The significance of this is twofold. Firstly, the application of Evidential Pluralism enables an evidence-based justification for legal interventions aimed at tackling online fake news. Secondly, the application of Evidential Pluralism in the context of online fake news provides a case study example to motivate a new methodology for evidence-based law, called EBL+.

The fact that Evidential Pluralism is able to overcome the difficulties in isolating and assessing the impact of online fake news and thereby provide an evidence-based justification for legal interventions aimed at addressing online fake news supports the use of

Evidential Pluralism in this case. Since most laws aim at addressing complex social issues, the difficulties involved in isolating and assessing the impact of online fake news will apply to the evaluation of other legal interventions and the problems they aim to address. The benefits of applying Evidential Pluralism to the case study example of online fake news therefore provides some motivation for EBL+, an approach to evidence-based law based on the principles of Evidential Pluralism.

The paper is structured as follows. Section 2 begins by clarifying that fake news is here being used as a convenient umbrella term to capture different kinds of false and misleading content. It then highlights the lack of evidence to support concerns over the detrimental impact of online fake news and explains some of the obstacles to gathering evidence of the impact of online fake news. Section 3 argues that evidence of the impact of online fake news is crucial to the justification of legal interventions to tackle online fake news. Section 4 sets out the core commitments of Evidential Pluralism. Section 5 provides an application of Evidential Pluralism to Covid-19 related online fake news to illustrate how combining evidence of correlation and evidence of mechanisms can provide robust overall evidence of the negative effects of online fake news. The fact that Evidential Pluralism enables an evaluation of the impact of online fake news and thereby an evidence-based justification for legal interventions supports the use of Evidential Pluralism in this case.

## 2. The (Assumed) Problem of Fake News

Fabricated and distorted information has long been used to try to influence and manipulate others to serve some personal or collective agenda. However, the internet, and especially

social media, facilitate the dissemination of so called 'fake news', leading to concerns it has become a problem like never before.[1]

'Fake news' is here being used as a convenient umbrella term to capture all kinds of false and misleading content.[2] In discussions of fake news, it is increasingly common to distinguish between misinformation, disinformation and mal-information.[3] Misinformation includes false or misleading content that is created or shared without an intention to do harm. Disinformation includes false or misleading content that is created or shared with the intention to deceive or manipulate. Mal-information includes content that is not false but is shared with the intention to manipulate or harm, such as publicly disseminating private messages or pictures.

These distinctions help to clarify the different kinds of false and misleading information and can be used to guide focused evidence gathering about the impact of specific kinds of problematic content (Wardle and Derakhshan, 2017; Southwell et al., 2019). Some kinds of

[1] For a detailed investigation into the ways in which computational technologies are utilized to spread false and misleading information, see Howard, (2020).

[2] Fake news is sometimes defined more narrowly to include only content that is intentionally deceptive and mimics news content. See, for example, Lazer et al., (2018). Understood in this more narrow way, fake news is a subtype of disinformation.

[3] For a detailed analysis of these concepts, see Wardle and Derakhshan, (2017). As Wardle and Derakhshan note, the term 'fake news' can itself be problematic or misleading because it has been utilized as a tool to distort or supress content that is not misleading or false. Politicians and powerful individuals have described content that is not misleading or false, but which conflicts with their interests or the narrative they are trying to promote, as 'fake news'. Nevertheless, the familiarity of the term and its convenience as a catch all phrase makes it useful for present purposes.

false or misleading content are unlikely to have much real-world significance. This could be because the content, although false, is not harmful. Alternatively, it could be because the harmful content fails to reach many people or fails to deceive those it does reach (Southwell et al., 2019). An understanding of which kinds of online fake news have a negative impact, and how, can guide which kinds of fake news, if any, should be subject to legal intervention. As we shall see, however, such evidence has, until recently, been lacking. In this paper, my concern is with the lack of evidence generally and the need to provide evidence to support legal interventions. I therefore do not narrow my focus to a specific kind of false or misleading content. 'Fake news' serves as a familiar and convenient umbrella term to capture the various different kinds of false and misleading information.[4]

Concerns over the detrimental effects of online fake news are widespread. In 2013, The World Economic Forum's Global Risks Report warned of 'digital wildfires' that could rapidly spread false information (World Economic Forum, 2013). A recent UK Government report claimed online disinformation threatens 'the very fabric of our democracy' (Digital, Culture, Media and Sport Committee, 2019). The World Health Organization (WHO) has labelled the spread of online fake news regarding the Covid-19 pandemic an 'infodemic', with WHO director-

[4] There are also issues with defining and categorizing different kinds of harms and which kinds of harm can justify legal interventions. While I do not deny these difficulties, I will not attempt to address them here. As we shall see in section 3, legal restrictions on content and its dissemination are constrained by freedom of expression and the ICCPR specifies protected interests that can justify legal restrictions. In my case study example, I rely on the ICCPR's specification of public health as a protected interest and the plausible claim that Covid-19 infection constitutes a harm and that protecting public health involves reducing the spread of Covid-19. See also footnote 13 for discussion of relevant harms.

general Tedros Adhanom Ghebreyesus saying, 'We're not just fighting a pandemic; we're fighting an infodemic' (World Health Organization, 2020; The Lancet Editorial, 2020).

Until recently, however, there has been little evidence provided to support the concern that fake news really does have a detrimental effect (Greene and Murphy, 2021; Lazer et al., 2018; Osman et al., 2023). For example, there is evidence that individual voters were regularly exposed to fake news during the 2016 US Presidential election and that much of the fake news was concentrated in highly competitive districts and targeted at potential Trump supporters (Allcott and Gentzkow, 2017; Howard, 2020; Howard et al., 2019). The fact that Trump ultimately won the election raised concerns that such fake news was effective at tipping the balance in Trump's favor. An observed correlation between fake news exposure and the election outcome is, however, insufficient to establish that the fake news actually influenced voter behavior and thereby the election outcome. Similarly, conspiracy theories about 5G masks being a cause of Covid-19 or making people more susceptible to severe infection by damaging their immune system were prevalent in the UK in the early stages of the pandemic. This led to concerns that arson attacks on 5G masks were driven by these conspiracy theories (BBC News, 2020). It might be thought that the specificity of the conspiracy claims and the associated arson attacks increases the likelihood that the observed correlation is causal. Further evidence would, however, increase confidence in the causal claim.[5] Furthermore, even if we accept the observed correlation as establishing causation, it applies to a very specific case of fake news and isolated events perpetrated by a minority of

[5] Further evidence has since been provided, see e.g. Jolley and Paterson, (2020).

individuals. Evidence that fake news has a more general detrimental effect is necessary to justify legal interventions.[6]

The lack of evidence of the detrimental impact of fake news is due, at least in part, to the fact that human behavior is a result of numerous, complex, interacting influences that make it difficult to isolate and determine the extent of the impact of fake news in a real-world context. Conducting robust randomized controlled trials (RCTs) could provide strong evidence of causation.[7] However, given concerns over the detrimental effects of fake news, purposefully exposing a large, representative cohort to fake news in the real-world social media environment would be unethical. Facebook staff and contract researchers previously published studies evidencing the extent to which they are able to influence people (Bond et al., 2012; Kramer, Guillory and Hancock, 2014). Concerns were raised over the unethical nature of the studies. As a result of this negative response, they no longer publish their findings (Howard, 2020).[8] RCTs conducted in controlled study environments can overcome ethical concerns but often fail to capture crucial aspects of the real-world environment, reducing their external validity (Howard, 2020). Observational studies can provide evidence that captures the real-world context. However, given the complexity of human behavior, it

[6] I return to this point in the following section.

[7] Standard evidence-based approaches prioritize meta-analyses and systematic reviews of RCTs and individual RCTS over other kinds of evidence. See, for example, OCEBM Levels of Evidence Working Group, (2011); What Works (2018).

[8] The fact that they no longer publish their findings does not mean that they no longer conduct relevant research. However, not publishing their work means (i) it cannot be assessed by others for scientific and ethical integrity, (ii) there are no external checks to ensure the research is being used to protect users, and (iii) the research cannot be used by others to understand the impacts of online fake news, including lawmakers.

can be difficult to control all variables in observational studies and therefore difficult to establish causation.

Despite these impediments to evidence gathering, concerns over the Covid-19 'infodemic' have prompted a move to gather evidence of the effect of Covid-19 related fake news on willingness to comply with public health measures, such as hand washing, wearing a face mask and social distancing (Allington et al., 2021; Roozenbeek et al., 2020), and rates of vaccine hesitancy and refusal (Greene and Murphy, 2021; Lockyer et al., 2021; Loomba et al., 2021; Neely et al., 2022; Pierri et al., 2022; Roozenbeek et al., 2020). Such public health measures are crucial to controlling the spread of Covid-19 and avoiding high numbers of severe illness and deaths and health services becoming overwhelmed (Gomes et al., 2022; Talic et al., 2021). Therefore, if Covid-19 related fake news reduces willingness to comply with public health measures, the impact could be significant, namely increased cases of severe illness, hospitalization and death from Covid-19.

Each of the studies referred to in the paragraph above provide some evidence of the effect of Covid-19 related online fake news on compliance with public health measures. Each individual study, however, does not provide sufficient evidence, on its own, that Covid-19 related online fake news has a significant, negative impact on public health. However, as I will show in section 5, an application of Evidential Pluralism enables us to identify and combine evidence from multiple studies to provide more robust, overall evidence for the concern that Covid-19 related online fake news has a real, detrimental impact. The fact that Evidential Pluralism is able to overcome the difficulties in evidence gathering and evaluate the impact of online fake news supports the use of Evidential Pluralism to evaluate the impact of online

fake news. In the following section, I argue that such evidence is crucial to the justification of legal interventions aimed at tackling the problem of online fake news.

## 3. The Importance of Evidence in Justifying Legal Interventions

Concerns over the detrimental effects of online fake news have prompted online service providers to implement self-regulated interventions, such as information correction and content removal policies.[9] Self-regulation of OSPs, however, is increasingly considered insufficient to protect users from harmful content (Woodhouse, 2020). As a result, numerous jurisdictions have proposed or implemented legislative or regulatory interventions to mitigate the negative impact of online fake news (Helm and Nasu, 2021). Such measures aim to ensure effective and consistent oversight of online content and its dissemination through legally enforceable standards and obligations.

The European Union's (EU) Digital Services Act (2022), for example, requires very large online platforms and search engines to conduct assessments of 'systemic risks' from the design and functioning of their services, including actual or foreseeable negative effects on civic discourse, electoral processes, public security and protection of public health. The assessments must consider the influence of intentional manipulation of their service and the 'amplification and potentially rapid and wide dissemination of illegal content and of information that is incompatible with their terms and conditions' (Digital Services Act 2022, Article 34(2)). In response to the risks identified, reasonable, proportionate and effective

[9] For example, see Facebook's Misinformation policy: Misinformation | Transparency Centre (fb.com) and Twitter's Synthetic and Manipulated Media Policy: Our synthetic and manipulated media policy | Twitter Help

mitigation measures must be implemented, such as adapting terms and conditions and content moderation processes (Digital Services Act 2022, Article 35(1)). After passing the Act, concerns over current practice and Elon Musk's publicly declared commitment to free speech led the EU to threaten to ban Twitter if it fails to comply with the content moderation requirements (Datta and D'Silva, 2022).

The legislative and regulatory interventions are driven by concerns over the negative impact of online fake news, including on public order, democratic stability and public health security. While these are important values to be protected, they must be balanced against competing interests. The fundamental normative issue arising from legal restrictions on online content and its dissemination is compatibility with freedom of expression.[10] Freedom of expression is considered an essential component of democratic societies and therefore holds a privileged and protected position. While restrictions on freedom of expression can be justified, they are typically subject to certain conditions. For example, the International Covenant on Civil and Political Rights (ICCPR), which includes the obligation to respect and ensure respect for freedom of expression, acknowledges that restrictions on any form of expression or means of its dissemination may be imposed if provided by law and *necessary* '(a) for respect for the rights or reputations of others; (b) for the protection of national security or of public order,

[10] This is not to say that there cannot be other relevant normative considerations. Freedom of expression is, however, a key normative consideration. The argument provided here for the need for evidence to justify restrictions on freedom of expression can be applied to other freedoms or interests. For discussion of concerns over the compatibility with freedom of expression of legal interventions aimed at tackling online fake news see, for example, Helm and Nasu (2021); Katsirea, (2018); Manzi, (2019).

or of public health or morals' (ICCPR, Article 19(3)).[11] Under the ICCPR, freedom of expression includes information and ideas that 'may shock, offend and disturb' and the falsity of information alone is not sufficient to justify restrictions (Human Rights Committee, 2011, at para 49).[12]

The protection of freedom of expression included in the ICCPR applies in the digital context. As the Human Rights Committee emphasizes, 'restrictions on the operation of websites, blogs or any other internet-based, electronic or other such information dissemination system, including systems to support such communication, such as Internet service providers or search engines, are only permissible to the extent that they are compatible with paragraph 3 [Article 19, ICCPR]' (Human Rights Committee, 2011, at para 43).

The condition of necessity specified in the ICCPR concerns the need for legislative or regulatory intervention and requires state parties to demonstrate the precise nature of the threat to a particular protected interest in a specific and individualized nature (Human Rights Committee, 2011, para 22 and 35). Concerns over the negative impact of online fake news include interests that, under the ICCPR, can justify restrictions on the content and dissemination of online fake news. However, while such concerns may seem reasonable, they do not seem sufficient to justify restrictions on freedom of expression in the absence of evidence to substantiate them (Katsirea, 2018; Marsden, Meyer and Brown, 2020). Surely,

[11] Non-signatories to the ICCPR include Malaysia, Myanmar and Singapore. These States have, however, signed the non-binding ASEAN Human Rights Declaration, which includes a right to freedom of expression and restrictions on free expression on the basis of protecting human rights, fundamental freedoms, national security, public order, public safety, public health, public morality and general welfare.

[12] See also, Organization for Security and Co-operation in Europe (2017).

'demonstrating the precise nature of the threat to a particular interest' requires evidence that online fake news *does* negatively impact protected interests such as national security or public health.[13] Furthermore, evidence of the kinds of fake news that cause a negative impact will help to ensure that restrictions are placed only on those kinds of fake news that have a negative impact and therefore help to avoid over-regulation.

Some minimally intrusive interventions might be understood as upholding freedom of expression and therefore thought to be justified without evidence of the negative impact of online fake news. Consider, for example, information correction interventions that provide access to accurate information and publicly debunk fake news but do not directly interfere with the content of or access to fake news. Such interventions can be understood as trying to improve the information available and thereby enhancing freedom of expression. Legal obligations to provide such information correction do, however, interfere with the freedom of expression of those on whom a legal obligation to provide information correction is being placed. Freedom of expression surely includes the freedom to refrain from expression,

[13] It might be thought that such restrictions could be justified on purely moral grounds. For example, it might be thought that intentional manipulation of others is wrong, in that it involves a failure to adequately respect those one is trying to manipulate. Such a perspective could be used to justify legal interventions to prevent such intentional manipulation without the need to demonstrate any further negative impact. It would still require normative justification that the harm of intentional manipulation outweighs freedom of expression, bearing in mind the ICCPR's protection of information that may 'shock, offend and disturb'. However, concerns over the negative impact of online fake news have not focused on such inherent wrongs. Instead, the focus has been on harms such as election manipulation and negative public health effects. Assuming that it is these kinds of harms with which the law is concerned, then evidence that such harms are occurring, or are likely to occur, is necessary to justify legal interventions.

including correcting others' false or misleading claims. More intrusive measures, such as content removal or criminal sanctions for creating and disseminating fake news, more clearly and significantly interfere with freedom of expression. Other measures, such as restrictions on data-driven, algorithmic micro-targeting, can still be understood to interfere with freedom of expression if algorithms are understood to fall under this right (Talia, 2022). Even if they are not, such measures do aim to control the flow of information and can therefore be understood as relevant to freedom of expression. Furthermore, such interventions place legal restrictions on the business practices of OSPs. Without evidence that online fake news has a negative impact, such restrictions do not seem justified.

Under the ICCPR, restrictions on freedom of expression are also subject to a test of proportionality that requires states to adopt the least intrusive intervention of those that might successfully protect against the identified threat (Human Rights Committee, 2011, para 34). Satisfying the proportionality test will require evidence of the effectiveness of specific legal interventions. But evidence of the negative impact of online fake news will also be important. If the negative impact of online fake news is only minimal, then more intrusive measures are unlikely to be proportionate, no matter how successfully they reduce the negative impact.

Concerns over the negative impact of online fake news are widespread. However, while such concerns may seem reasonable, they do not seem sufficient to justify legal interventions in the absence of evidence to substantiate them. Restrictions on the content and dissemination of online content must be compatible with freedom of expression and this requires demonstrating that they are both necessary and proportionate. Providing evidence of the negative impact of online fake news seems crucial to satisfying both the necessity and

proportionality requirements. Without such evidence, legal interventions risk lacking justification. The remainder of the paper offers a framework for providing such evidence.

## 4. Evidential Pluralism

Evidential Pluralism is a philosophical account of causal enquiry.[14] Evidential Pluralism includes the following two pluralist principles (Shan and Williamson, 2023):

*Object Pluralism*: In order to establish a causal claim, one normally needs to establish two propositions: (i) that the putative cause and effect are appropriately correlated, and (ii) that there is some mechanism complex involving the putative cause which is responsible for the putative effect, and which can account for the extent of the correlation.

*Study Pluralism:* In order to assess a causal claim, one normally needs to assess relevant association studies and mechanistic studies, where available.

A key motivation for object pluralism is the truism that correlation is not causation. That A is a cause of B is one possible explanation for an observed correlation between A and B. There are, however, many other possible explanations. The observed correlation might, for example, be the result of B causing A. Alternatively, it might be due to confounding, bias or chance. When a correlation between A and B is the result of A causing B, there is some mechanism complex by which A leads to B and that can account for the extent of the correlation. Therefore, in order to establish that A is a cause of B, one must establish

[14] Evidential Pluralism was originally proposed by Russo and Williamson (2007). It has since been developed, defended and applied, including by: Auker-Howlett and Wilde, (2019); Illari (2011); Illari and Williamson (2012); Parkkinen, et al., (2018); Shan and Williamson (2021); Shan (2022); Shan and Williamson (2023).

mechanism as well as correlation. Further motivation for object pluralism comes from the fact that causal claims have two key objectives, (i) diagnosing, predicting and intervening, and (ii) explanation. For diagnosis, prediction and intervention, a cause must make a difference to its effects, and therefore establishing correlation is crucial. On the other hand, mechanisms are crucial to explanation. Therefore, to satisfy the two key objectives of causal claims, it is necessary to establish both correlation and mechanisms.[15]

Study pluralism follows from object pluralism. Since establishing causation requires establishing both correlation and mechanism and establishing mechanism will normally require assessing studies other than those establishing correlation, it is normally necessary to assess both 'association studies' and 'mechanistic studies'.

Association studies, according to Evidential Pluralism, establish correlation. Establishing a correlation between A and B normally involves carrying out an experimental or observational study that repeatedly measures instances of A and B, together with potential confounders, to determine whether A and B are probabilistically dependent, conditional on the measured confounders. Association studies include, for example, RCTs, cohort studies and case control studies.

RCTs are viewed as providing particularly strong evidence of causation because they help to reduce the probability that a correlation is attributable to unforeseen confounders, thereby raising the probability that there is some mechanism responsible for the correlation. In some cases, sufficiently robust RCTs can provide direct evidence of correlation and indirect

[15] These are the motivations for Evidential Pluralism originally put forward in Russo and Williamson (2007). For further support for the claim that mechanisms are crucial to explanation, see Salmon (1998) and Machamer, Darden and Craver, (2000).

evidence of mechanisms, and thereby establish causation via channels α1 and 2 (Fig 1). In many cases, however, RCTs are insufficiently robust to establish mechanism indirectly. In such cases, the mechanism claim can be investigated directly by assessing mechanistic studies.

According to Evidential pluralism, investigation of the mechanism begins by hypothesizing specific features of the mechanism complex linking A and B. Studies that test for the presence of these features are classified by Evidential Pluralism as mechanistic studies. On Evidential Pluralism, a mechanism consists of a structured arrangement of parts, including intermediary variables, entities, activities or processes, or some combination of these. To be confident that an identified mechanism does account for the phenomenon to be explained, it is important to consider the whole complex of mechanisms linking A and B, including any reinforcing or counteracting mechanisms. Evidential Pluralism does not, however, require all the details of a mechanism to be established. Provided a specific mechanism hypothesis that posits key features of the mechanism linking A and B is established and accounts for the extent of the observed correlation between A and B, confirmation of the causal claim can be provided.[16]

[16] For further explanation of how Evidential Pluralism understands mechanisms, see Illari and Williamson (2012) and Shan and Williamson (2023), pp. 8-10 and p. 19. See also Glennan and Illari (2018) for the philosophical foundations on which Evidential Pluralism's understanding of mechanisms draws.

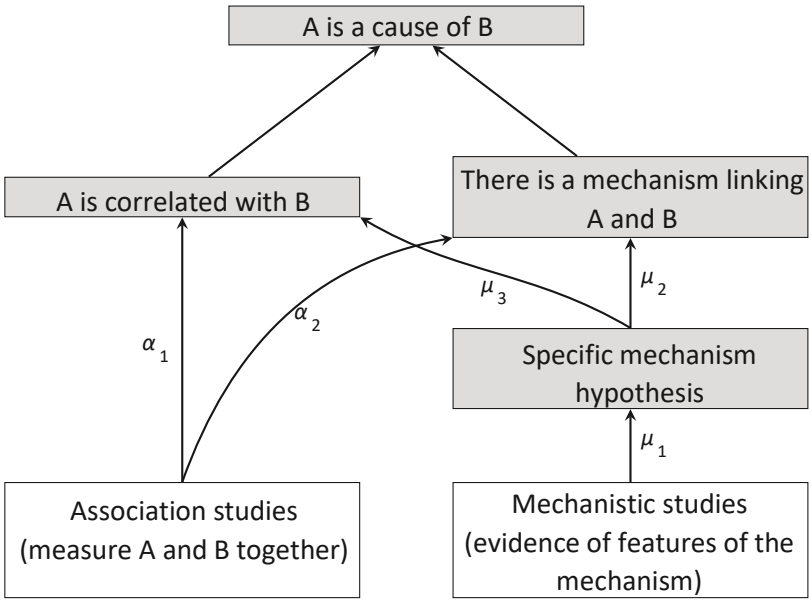


Figure1. Evidential Pluralism: Evidential Relationships for Establishing Causation (Shan and Williamson, 2023)[17]

The requirement to assess mechanistic studies alongside association studies motivates a departure from standard evidence-based approaches that prioritize RCTs and observational studies, including those of evidence-based medicine (EBM) and evidence-based policy (EBP). Evidential Pluralism has enabled the development of a more inclusive approach to evidence evaluation in medicine, EBM+ (Parkkinen et al., 2018), and policy, EBP+ (Shan and Williamson, 2023). A similar application to law motivates a new methodology for evidence-based law, EBL+.[18]

[17] Reproduced under CC-BY-NC-ND License.

[18] For a more detailed explanation of the application of Evidential Pluralism to evidence-based law, see (Trofimov and Williamson, forthcoming).

In this paper, I am relying on the general motivation for Evidential Pluralism briefly outlined above and developed and defended in previous work by others. Building on this foundation, I provide practical motivation for the use of Evidential Pluralism to evaluate the impact of online fake news, which in turn provides some motivation for the adoption of Evidential Pluralism as a methodology for evidence-based law more generally.

The problem of online fake news illustrates the need for a more inclusive approach to evidence evaluation in evidence-based law. As discussed in section 2, the complexity of the problem means that RCTs and observational studies will not be sufficient to evaluate the impact of online fake news. As we shall see in the following section, however, Evidential Pluralism enables us to overcome these difficulties. By providing a methodology for systematically combining different kinds of evidence, Evidential Pluralism enables an evaluation of the impact of online fake news. This, in turn, enables an evidence-based justification for legal interventions aimed at addressing the problem of online fake news. The fact that Evidential Pluralism is able to overcome the difficulties in evaluating the impact of online fake news and thereby enable an evidence-based justification for legal interventions supports the use of Evidential Pluralism in this case.

The benefits of applying Evidential Pluralism to the problem of online fake news provides a case study example to motivate the adoption of Evidential Pluralism as a methodology for evidence-based law more generally. Since most laws aim to address complex social issues, the difficulties involved in conducting and relying on RCTs and observational studies in the case of online fake news will arise when evaluating other legal interventions and the problems they aim to address. As the example of online fake news illustrates, Evidential Pluralism can overcome these difficulties. A full defence of Evidential Pluralism as a methodology for

evidence-based law will require development of the theory of EBL+ and consideration of further case studies. Nevertheless, the benefits of applying Evidential Pluralism to the problem of online fake news provides some motivation for EBL+.

The remainder of the paper focuses on demonstrating the benefits of EBL+ through an application of Evidential Pluralism to the problem of Covid-19 online fake news.

## 5. EBL+: Applying Evidential Pluralism to Support the Concern over Fake News

Evidential Pluralism offers a methodology for systematically combining evidence to enable an evaluation of the impact of online fake news and thereby enable an evidence-based justification for legal interventions aimed at addressing the problem of online fake news.

According to Evidential Pluralism, to establish the causal claim that online fake news has a negative impact, both correlation and mechanism must be established. Given the limitations to conducting RCTs to assess the real-world impact of fake news discussed above in section 2, establishing both correlation and mechanism through sufficiently robust RCTs is not feasible (via α1 and α2, Fig 1). However, establishing correlation through association studies and establishing mechanism directly through the assessment of mechanistic studies is feasible (via α1 and μ1, Fig 1). Evidential Pluralism, therefore, enables us to assess the real-world effects of online fake news by combining evidence of correlation and evidence of mechanisms. The fact that Evidential Pluralism enables an evaluation of the impact of online fake news supports the use of Evidential Pluralism to evaluate online fake news.

To illustrate the advantages of Evidential Pluralism, I here provide an example of the methodology of applying Evidential Pluralism to assess the impact of Covid-19 related fake

news on compliance with public health measures. In providing this example, my aim is to (i) clarify how an Evidential Pluralism evaluation proceeds, (ii) show that an Evidential Pluralism evaluation is feasible, and (iii) support the claim that Evidential Pluralism enables an evaluation of the impact of online fake news.

### 5.1: The Causal Claim

An Evidential Pluralism evaluation begins by specifying the causal claim of interest. In this case, the causal claim of interest is that Covid-19 related online fake news causes a decrease in compliance with crucial public health measures, such as social distancing and accepting a vaccination. Establishing this causal claim, according to Evidential Pluralism, requires establishing both correlation and mechanisms.

### 5.2: Evidence of Correlation

The correlation claim of interest is that Covid-19 related online fake news is negatively correlated with compliance with public health measures, conditional on potential confounders.

Studies that provide evidence of a correlation between Covid-19 related online fake news and public health behaviors have been conducted. For example, Neely et al., (2022) conducted a web-based survey of a representative sample of 600 adults in the state of Florida, U.S.A., in June 2021. They found high levels of misinformation exposure, with 73.2% of participants reporting some exposure to Covid-19 vaccine misinformation in the previous 6 months. A 'statistically significant correlation' between vaccination status and exposure to misinformation was observed (Neely et al., 2022, p. 182). Vaccination levels were 73.8% among participants who reported no exposure to misinformation, 62.9% among participants

who reported exposure to one misinformation theme and 52.2% among participants who reported exposure to six or more misinformation themes. Misinformation exposure was found to be a strong predictor of vaccine hesitancy, conditional on other demographic and political predictors of vaccine hesitancy.[19]

Pierri et al. (2022) also provide evidence of correlation between misinformation exposure and vaccination status across 50 U.S. states. They investigated the relationship between the mean percentage of misinformation shared via Twitter at State level and State level vaccination rates. They found that an increase in the mean amount of online misinformation shared is significantly associated with a decrease in daily vaccination rates per million, conditional on potential confounders known to be associated with vaccine uptake or hesitancy.

These studies provide evidence of correlation between misinformation and vaccine rates and control for a number of confounders. However, it is possible that some bias or confounding persists. Furthermore, the studies look only at vaccine rates and not other public health interventions. Therefore, the studies are not sufficient on their own to establish the claim that Covid-19 related fake news has a negative impact on compliance with public health measures. Evidence of mechanisms can be used to supplement these results to provide more robust overall evidence that Covid-19 fake news has a behavioral impact.

### 5.3: Evidence of Mechanisms

According to Evidential Pluralism, investigation of mechanisms begins by providing a mechanism hypothesis. This hypothesis can then be used to guide evidence gathering and identify any gaps in existing evidence. If existing evidence establishes a feature of the

[19] Political affiliation was also found to be a strong predictor.

mechanism, then there is no need to gather further evidence of that feature. However, if any key feature of the mechanism is not yet established, then further evidence gathering is required.

A plausible hypothesis of the mechanism connecting fake news and behavior is that exposure to fake news impacts individual beliefs which impact individual behavioral intentions which impact individual behavior:

Fake News Exposure → Beliefs → Behavioral Intentions → Behavior

Figure 2. Mechanism Hypothesis

For each stage of this mechanism hypothesis, there are studies that provide supporting evidence.

#### 5.3.1: Behavioral Intentions → Behavior

Let's begin with the link between behavioral intentions and behavior. The idea that behavioral intentions are a predictor of behavior is a central commitment of many social psychological models of behavior. Sheeran (2002) conducted a meta-analysis of meta-analyses of the behavior-intention relationship to see how well behavioral intentions predict behavior. Sheeran found a sample-weighted average correlation of 0.53 between intentions and behavior, meaning that, on average, intentions account for twenty eight per cent of the variance in behavior. While this does not seem to be a substantial correlation, Sheeran explains that difficulties in interpreting percentage variance result in such values appearing much less significant than they are. Drawing on Cohen (1992) as a guide to interpreting the

correlations, Sheeran concludes that 'explaining 28% of the variance should probably be considered 'good'' (2002, pp. 4-5). Furthermore, Sheeran found that those who fail to act on a positive intention to act ('inclined abstainers') account for the majority of the identified gap or inconsistency between behavioral intentions and behavior, whereas those who did not intend to act but subsequently did so ('disinclined actors') accounted for only a very small percentage. This supports the idea that while various factors, such as strength or type of intention, type of behavior or perceived behavioral control, can result in behavioral intentions not resulting in the intended behavior, it is unlikely that a person will act in the absence of an intention to do so.[20] This is important for present purposes. If fake news exposure results in individuals failing to form an intention to engage in Covid protective behaviors, such as receiving a vaccine or complying with social distancing measures, then the chances of them actually engaging in such behavior is minimal. On the other hand, if fake news exposure is shown to have little or no impact on behavioral intentions, then there is little chance of it impacting behavior (Greene and Murphy, 2021).

#### 5.3.2: Beliefs → Behavioral Intentions

Turning to the link between beliefs and behavioral intentions, there are studies that provide evidence that belief in Covid-19 related fake news influences behavioral intentions. For example, Roozenbeek et al. (2020) conducted large national surveys in the UK, Ireland, USA, Spain and Mexico to investigate the influence of susceptibility to Covid-19 related misinformation on willingness to receive and recommend vaccination and willingness to

[20] Intentions are not here being attributed wherever there is action as part of a folk psychological explanation of behavior. Furthermore, it is not assumed that all intentions are conscious. For a more detailed consideration of behavioral intentions, see Sheeran, (2002).

comply with public health measures. Susceptibility to misinformation was measured by asking participants to rate the reliability of common Covid-19 misinformation statements. Statistical analysis of the survey data found (i) belief in one conspiracy strongly correlates with belief in others, (ii) being exposed to misinformation online is associated with higher susceptibility to misinformation and (iii) higher susceptibility to misinformation is associated with a decrease in reported willingness to receive and recommend vaccination and a decrease in reported willingness to comply with public health measures. Higher susceptibility to misinformation was the only variable that predicts lower compliance with public health guidance. The robust associations across cultural contexts in large national samples found in this study support the idea that belief in misinformation influences behavioral intentions, an important stage in the proposed mechanism above.

#### 5.3.3: Beliefs → Behavioral Intentions → Behavior

There are also studies that provide evidence that belief in Covid-19 fake news influences not only behavioral intentions but also behavior. For example, Romer and Jamieson (2020) conducted a national probability survey with a representative sample of 1050 US adults in March 2020 and a follow in July 2020 with 840 of the same individuals. The study investigated the relationship between belief in Covid-19 related conspiracy theories and taking preventive actions and vaccination intentions. They controlled for a number of variables, including political ideology, demographics and news consumption. They found that belief in Covid-19 related conspiracy theories was negatively correlated with (i) perceived threat of the pandemic, (ii) taking preventive actions, (iii) perceived safety of vaccination, and (iv) intentions to be vaccinated. The follow up survey enabled analysis of (i) the stability of conspiracy beliefs, (ii) whether conspiracy beliefs held in March 2020 were negatively

correlated with preventive actions and vaccination intentions in July 2020, and (iii) whether conspiracy beliefs held in March 2020 were predictive of subsequently recommended preventive actions such as mask wearing. They found that conspiracy beliefs were stable and that conspiracy beliefs in March 2020 predicted vaccination intentions and preventive actions in July 2020, including subsequently recommended mask wearing.

#### 5.3.4: Fake News Exposure → Behavioral Intentions[21]

Other studies provide evidence of the link between exposure to online fake news and behavioral intentions. For example, Loomba et al. (2021) conducted a RCT involving 8,001 participants across the USA and the UK to measure the effect of Covid-19 related online misinformation on vaccination intentions. Participants were asked about their intention to receive a Covid-19 vaccination. Participants in the treatment group were then exposed to Covid-19 related misinformation while participants in the control group were exposed to factual information. Intentions to accept a vaccination were then re-assessed. The study found exposure to misinformation concerning Covid-19 vaccinations lowers individuals' intentions to receive a vaccination to protect themselves and lowers their intentions to receive a vaccination to protect others. They found, for example, that misinformation decreases the number of respondents who would 'definitely' accept a vaccine relative to the control group by 6.2% in the UK and 6.4% in the USA. Furthermore, for any pre-treatment response, exposure to misinformation results in a net movement towards the response category immediately below. For example, among UK treatment participants who responded

[21] Although this is not a direct link in the mechanism hypothesis, exposure to fake news and behavioral intentions are intermediaries and therefore evidence that supports a connection between them is still relevant mechanistic evidence.

pre-exposure ‘unsure, but leaning towards yes’, they found a 10.6% increase in the post-exposure response ‘unsure, but leaning towards no’.

#### 5.3.5: Exposure → Beliefs → Behavioral Intentions

The study by Loomba et al. provides evidence of the mechanism connecting exposure to misinformation and behavioral intentions. A study by Greene and Murphy (2021) similarly provides evidence of this mechanistic stage but also provides evidence of the intermediary belief stage. Greene and Murphy investigated the effects of a one-off exposure to Covid-19 misinformation on behavioral intentions. To avoid previous exposure confounding results, participants were exposed to novel Covid-19 related misinformation stories that were fabricated by the study authors and had not previously been reported in the media. Each fabricated story related to one of the behaviors under investigation and each participant was exposed to two of the fabricated stories. Small but significant changes to two of the four behavioral intentions investigated were associated with exposure to misinformation. The observed effects represented a change of approximately 5%-12% in behavioral intention relative to participants not exposed to the fake news story. In addition to assessing behavioral intentions, Greene and Murphy asked participants to rate the truthfulness of the four fake stories and found that participants who believed the fake news stories to be truthful reported stronger intentions to engage in the behavior suggested by the story. These findings support the idea that the relationship between exposure to misinformation and behavioral intentions is mediated by belief in the misinformation.

Although the studies by Loomba et al and Greene and Murphy provide evidence of the mechanism connecting exposure to misinformation and behavioral intentions, the associations observed in each study are relatively small. Of course, even small changes in

individual behavior can have significant social effects. As Greene and Murphy highlight, following misinformation linking the measles, mumps and rubella (MMR) vaccination with autism, there was a drop in childhood MMR vaccination rates in the early 2000s which led to a reduction in herd immunity and an increase in cases of measles.[22] Furthermore, as the authors of both studies acknowledge, the study environments do not replicate crucial aspects of the real-world environment. The results are therefore unlikely to be representative of the effect of misinformation in a real-world social media context. Having evidence of the more fine-grained mechanistic stages of the relationship between exposure to fake news and behavior can help to evaluate the extent to which the results from these studies extrapolate into a real-world context.

#### 5.3.6: Exposure → Beliefs

There are various aspects of the real-world online environment not captured in the studies by Loomba et al. (2021) and Greene and Murphy (2021) that have been shown to be important in the relationship between exposure to fake news and beliefs. For example, repeated exposure to fake news is an aspect of the real-world social media environment not captured in the studies by Loomba et al. and Green and Murphy. Pennycook et al. (2018) found repeated exposure can increase perceived truthfulness of fake news, even after just two exposures.[23] Sharing information amongst trusted networks of friends and family is another element of the real-world social media environment not replicated in the studies by Loomba et al. (2021) and Greene and Murphy (2021). Allington et al. (2021) found a 'small

[22] See also Leask, Booy and McIntyre (2010).

[23] See also Fazio (2016); Fazio, et al., (2015).

but significant positive relation' between belief in one or more Covid-19 conspiracy theories and using friends and family as a source of information on Covid-19. They also found a weak but 'still significant negative relationship' between using friends and family as a source of knowledge about Covid-19 and engagement in health protective behaviors. The results of these studies provide evidence that the impact of exposure to fake news on beliefs and in turn behavior in the real-world social media environment would be more significant and severe than those identified in the studies by Loomba et al. (2021) and Greene and Murphy (2021).

Research on psychological mechanisms such as confirmation bias provide further insights into the link between fake news exposure and belief.[24] Confirmation bias involves seeking out and believing information that is consistent with or supportive of one's existing beliefs, values and attitudes and to interpret information in line with these existing beliefs, attitudes and values. Disconfirmation bias involves being more critical of and discounting information that is inconsistent with one's existing beliefs, attitudes and values. These psychological mechanisms have also been shown to involve a tendency to consider a source more credible and trustworthy when content is consistent with and supportive of existing beliefs, attitudes and values and less credible and trustworthy when content is inconsistent with or critical of existing beliefs, attitudes and values (Bastardi, Uhlmann and Ross, 2011; Beauvais, 2022; Benegal and Scruggs, 2018; Flynn, Nyhan and Reifler, 2017; Kahan et al., 2010; Kahan, Jenkins-Smith and Braman, 2011; Pennycook and Rand, 2018; Westerwick, Kleinnman and Knobloch-

[24] Helm and Nasu (2021) draw on empirical evidence relating to these psychological mechanisms to evaluate legal interventions aimed at tackling online fake news. They argue that information correction and content blocking are unlikely to be successful given these psychological mechanisms.

Westerwick, 2013). Engaging with and sharing fake news that aligns with existing beliefs, attitudes and values facilitates the spread of fake news, enabling greater exposure, especially amongst like-minded individuals, while being disposed to accept information that aligns with existing beliefs, attitudes and values facilitates belief in fake news (Ecker et al. 2022; Pennycook and Rand, 2021).

### 5.4: Evaluation Conclusion

On the evidence considered here, each link in the mechanism hypothesis has, plausibly, been established. The combination of evidence of correlation and evidence of mechanisms considered here plausibly establishes that Covid-19 related fake news does negatively impact compliance with crucial public health measures. This establishes that Covid-19 related fake news is a problem that needs to be addressed and thereby helps to justify legal interventions aimed at tackling the negative impact of Covid-19 related fake news.

It is important to stress that the aim of this section has been to show how an application of Evidential Pluralism overcomes the limits to establishing difference making in complex human behavior and thereby enables an evaluation of the impact of Covid-19 related online fake news. A full and systematic review of the evidence has not been conducted and therefore the conclusions are tentative. Nevertheless, the results demonstrate the benefits of combining evidence of correlation and evidence of mechanisms and suggest that a thorough, systematic review of the evidence would establish that Covid-19 related fake news has a negative impact on compliance with crucial public health measures.

It is also important to stress that I am not arguing that legal interventions ought to be implemented to address the problem of online fake news. As I argued in section 3, evidence that online fake news has a negative impact on protected interests is crucial to the

justification of legal interventions aimed at addressing online fake news. Restrictions on online content and its dissemination must be compatible with freedom of expression. This requires demonstrating that they are both necessary and proportionate which in turn requires evidence of the negative impact of online fake news. Even once the negative impact of online fake news is established, however, judgements must be made about whether the extent of the negative impact is sufficient to justify restrictions on freedom of expression. Furthermore, evidence of the effectiveness of specific legal interventions will be required to make a judgement about the proportionality of specific legal interventions. I am not providing such judgements. Nevertheless, evidence of the impact of online fake news is crucial to satisfying both necessity and proportionality and therefore to justifying legal interventions.

The fact that Evidential Pluralism provides a framework that enables the impact of online fake news to be evaluated and thereby enable an evidence-based justification of legal interventions supports the use of Evidential Pluralism as a methodology for evaluating the impact of online fake news.

## 6. Conclusion

The aim of this paper has been to argue that Evidential Pluralism ought to be used to evaluate the impact of online fake news to enable an evidence-based justification for legal interventions aimed at addressing the problem of online fake news.

Evidential Pluralism's requirement to assess evidence of mechanisms alongside evidence of correlation overcomes the difficulties in evaluating the impact of online fake news. The fact

that Evidential Pluralism provides a framework for systematically combining evidence to enable an evaluation of the impact of online fake news supports its use in this case.

The benefits of Evidential Pluralism have been illustrated through an application to Covid-19 related fake news. The combination of evidence of correlation and evidence of mechanisms considered plausibly establishes that Covid-19 related fake news does have a negative impact on compliance with crucial public health measures. Given the importance of compliance with public health measures in controlling the spread of Covid-19, this evidence can be used to justify the concern that Covid-19 related fake news is a problem that needs to be addressed. Establishing that Covid-19 related fake news has a negative impact makes it likely that other kinds of fake news will also have a negative impact. An application of Evidential Pluralism to other kinds of fake news could help to establish which other kinds have a negative impact.

Evidence that online fake news has a negative impact is crucial to the justification of legal interventions aimed at mitigating the impact of online fake news. Legal restrictions on online content and its dissemination must be compatible with freedom of expression, which requires demonstrating that legal interventions are necessary and proportionate. While concerns over the negative impact of online fake news seem reasonable, they must be substantiated by evidence to justify legal interventions.

Evidential Pluralism provides a framework that enables an evidence-based justification for legal interventions aimed at mitigating the negative effects of online fake news. The benefits of applying Evidential Pluralism to the problem of online fake news provides a case study example to motivate the adoption of Evidential Pluralism as a new methodology for evidence-based law, EBL+. To make the case that Evidential Pluralism applies more generally and should

be adopted as an evidential framework for evidence-based law, it is important to consider further case studies and develop the theory of EBL+. This will be the focus of future research.[25]

Acknowledgements: Thanks to Jon Williamson for detailed feedback on earlier drafts of this paper. Earlier versions of this work were presented at Østfold University College, Norway, University College London and the University of Kent. Thanks to participants of those events for helpful discussions.

Funding: Alexandra receives funding from Leverhulme Trust, Grant RPG-2022-336.

References

Allcott, H. and Gentzkow, M., (2017), 'Social Media and Fake News in the 2016 Election', *Journal of Economic Perspectives*, 31 (2): 211-36.

Allington, D., Duffy, B., Wessely, S., Dhavan, N. and Rubin, J., (2021), 'Health-Protective Behaviour, Social Media Usage and Conspiracy Belief During the COVID-19 Public Health Emergency', ***Psychological Medicine,* 51(10): 1763-1769.**

Auker-Howlett, D. and Wilde, M., (2019), 'Reinforced Reasoning in Medicine', *Journal of Evaluation in Clinical Practice* 26(2) 458-464.

Bastardi, A., Uhlmann, E. L. and Ross, L., (2011), 'Wishful Thinking: Belief, Desire and the Motivational Evaluation of Scientific Evidence', *Psychological Science, 22(6): 731-732.*

[25] See, for example, (Trofimov and Williamson, forthcoming).

BBC News, (2020), 'Ofcom: Covid-19 5G Theories are 'Most Common' Misinformation'. Available at: Ofcom: Covid-19 5G theories are 'most common' misinformation - BBC News

Beauvais, C., (2022), 'Fake News: Why do we Believe it?', *Joint Bone Spine,* 89(4).

Benegal, S. D. and Scruggs, L., (2018), 'Correcting Misinformation about Climate Change: The Impact of Partisanship in an Experimental Setting', *Climatic Change,* 148(3).

Bond, R. M., Fariss, C. J., Jones, J. J., Kramer, A. D., Marlow, C., Settle, J. E., Fowler, J. H., (2012), 'A 61-Million-Person Experiment in Social Influence and Political Mobilization', *Nature*, 489(7415): 295-8.

Cohen, J., (1992), 'A Power Primer', *Psychological Bulletin*, 112: 155-159.

Datta, T. and D'Silva, A., 'EU Warns Musk that Twitter Faces Ban Over Content Moderation – FT', *Reuters*. Available at: EU warns Musk that Twitter faces ban over content moderation -FT | Reuters

Digital, Culture, Media and Sport Committee, (2019), 'Disinformation and Fake News: Final Report', Eighth report ofSession 2017-2019, HC 1791, House of Commons, UK Parliament. Available at: Disinformation and 'fake news' (parliament.uk)

Ecker, U. K. H., Lewandowsky, S., Cook, J., Schmid, P., Fazio, L. K., Brashier, N., Kendeou, P., Vraga, E. K. and Amazeen, M. A., (2022), 'The Psychological Drivers of Misinformation Belief and its Resistance to correction', *Nature Reviews Psychology*, 1: 13-29.

Fazio, L., (2016), 'Unbelievable News? Read it Again and you might Think It's True', *The Conversation*. Available at: Unbelievable news? Read it again and you might think it's true (theconversation.com)

Fazio, L., Brashier, N. M., Payne, B. K. and Marsh, E. J., (2015), 'Knowledge Does Not Protect Against Illusory Truth', *Journal of Experimental Psychology: General*, 144(5): 993-1002.

Flynn, D. J., Nyhan, B. and Reifler, J., (2017), 'The Nature and Origins of Misperception: Understanding False and Unsupported Beliefs about Politics', *Advances in Political Psychology,* 38(1): 127-150.

Glennan, S. and Illari, P, (eds.) (2018), *The Routledge Handbook of Mechanisms and Mechanistic Philosophy*, Routledge.

Gomes, M. G. M., Ferreira, M. U., Corder, R. M., King, J. G., Souto-Maior, C., Penha-Goncalves, C., Chikina, M., Pegden, W. and Aguas, R., (2022), 'Individual Variation in Susceptibility or Exposure to SARS-CoV-2 Lowers the Herd Immunity Threshold', *Journal of Theoretical Biology*, 540.

Greene, C. M. and Murphy, G., (2021), 'Quantifying the effects of fake news on behavior: Evidence from a study of Covid-19 Misinformation', *Journal of Experimental Psychology: Applied,* 27(4): 773-784

Helm, R. K. and Nasu, H., (2021), 'Regulatory Responses to Fake News and Freedom of Expression: Normative and Empirical Evaluation', *Human Rights Law Review,* 21: 302-328

Howard, P. N., (2020), *Lie Machines,* Yale University Press.

Howard, P. N., Neudert, L-M., Bradshaw, S. and Kollanyi, B., (2019), 'Sourcing and Automation of Political News and Information over Social Media in the Unites States, 2016-2018', *Political Communication*, 37(2): 173-193.

Human Rights Committee, (2011), General Comment 34: Freedom of Expression (art. 19). Available at: General comment No.34 on Article 19: Freedoms of opinion and expression | OHCHR

Illari, P. M., (2011), 'Mechanistic Evidence: Disambiguating the Russo-Williamson Thesis', *International Studies in the Philosophy of Science,* 25(2): 139-157

Illari, P. M. and Williamson, J., (2012), 'What is a Mechanism? Thinking about Mechanisms Across the Sciences', *European Journal for Philosophy of Science*, 2: 119-135.

Jolley, D., and Paterson, J. L., (2020), 'Pylons Ablaze: Examining the Role of 5G Covid-19 Conspiracy Beliefs and Support for Violence', *British Journal of Social Psychology*, 59(3): 628-640.

Kahan, D. M., Braman, D., Cohen, G. L., Slovic, P. and Gastil, J., (2010), 'Who Fears the HPV Vaccine, Who Doesn't and Why? An Experimental Study of the Mechanisms of Cultural Cognition', *Law and Human Behaviour,* 34: 501-516

Kahan, D. M., Jenkins-Smith, H. and Braman, D., (2011), 'Cultural Cognition of Scientific Consensus', *Journal of Risk Research,* 14(2): 147-174.

Katsirea, I., (2018), '"Fake News": Reconsidering the Value of Untruthful Expression in the Face of Regulatory Uncertainty', *Journal of Media Law*, 10(2): 159-188.

Kramer, A. D., Guillory, J. E., Hancock, J. T., (2014), 'Experimental Evidence of Massive-Scale Emotional Contagion through Social Networks. *Proceedings of the National Academy of Sciences of the USA*, 111(24): 8788-90.

Lazer, D. M., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., Metzger, M J., Nyhan, B., Pennycook, G., Rothshild, D., Schudson, M., Sloman, S. A., Sustein, C. R., Thorson, E. A., Watts, D. J., Zittrain, J. L., (2018), 'The Science of Fake News', *Science*, 359(6380): 1094-1096.

Leask, J., Booy, R. and McIntyre, P. B., (2010), 'MMR, Wakefield and The Lancet: What can we learn?', *The Medical Journal of Australia,* 193(1), 5–7

Lockyer, B., Islam, S., Rahman, A., Dickerson, J., Pickett, K., Sheldon, T., Wright, J., McEachan, R., Sheard, L., (2021), 'Understanding Covid-19 Misinformation and Vaccine Hesitancy in Context: Findings from a Qualitative Study involving Citizens in Bradford, UK,' *Health Expectations*, 24(4): 1158-1167.

Loomba, S., de Figueiredo, A., Piatek, S. J. de Graaf, K. & Larson, H. J. (2021) Measuring the impact of COVID-19 vaccine misinformation on vaccination intent in the UK and USA. *Nature Human Behaviour,* 5: 337-348.

Machamer, P. Darden, L. and Craver, C., (2000), 'Thinking About Mechanisms', Philosophy of Science, 67(1): 1-25.

Manzi, D. C., (2019), 'Managing the Misinformation Marketplace: The First Amendment and the Fight Against Fake News' Fordham Law Review, 87(6): 2623.

Marsden, C., Meyer, T., Brown, I., (2020), 'Platform Values and Democratic Elections: How can the Law Regulate Digital Disinformation?', *Computer Law and Society*, 36: 105373

Neely, S. R., Eldredge, C., Ersing, R. and Remington, C., (2022), Vaccine Hesitancy and Exposure to Misinformation: A Survey Analysis, *Journal of General Internal Medicine,* 37(1): 179-187.

OCEBM Levels of Evidence Working Group, (2011), 'The Oxford 2011 Levels of Evidence', Oxford Centre for Evidence-Based Medicine. Available at: OCEBM Levels of Evidence — Centre for Evidence-Based Medicine (CEBM), University of Oxford

Organization for Security and Co-operation in Europe, (2017), 'Joint Declaration on Freedom of Expression and 'Fake News', Disinformation and Propaganda'. Available at: Joint declaration on freedom of expression and "fake news", disinformation and propaganda | OSCE

Osman, M., Meder, B., Bechlivanidis, C. and Adams, Z., (2023), 'Misinformation: Why it may not lead to bad behaviour' *The Conversation.* Available at: Misinformation: why it may not necessarily lead to bad behaviour (theconversation.com)

Parkkinen, V-P., Wallmann, C., Wilde, M., Clarke, B., Illari, P., Kelly, M. P., Norell, C., Russo, F., Shaw, B., and Williamson, J., (2018), *Evaluating Evidence of Mechanisms in Medicine: Principles and Procedures,* Springer, Cham, Switzerland.

Pennycook, G., Cannon, T., Rand, D.G., (2018), 'Prior Exposure Increases Perceived Accuracy of Fake News'**,** *Journal of Experimental Psychology General,* 147(12)**:** 1865-1880.

Pennycook, G. and Rand, D. G., (2018), 'Lazy, Not Biased: Susceptibility to Partisan Fake News is Better Explained by Lack of Reasoning than by Motivated Reasoning', *Cognition,* 188: 39-50.

Pennycook, G. and Rand, D. G., (2021), 'The Psychology of Fake News', *Trends in Cognitive Sciences,* 25(5).

Pierri, F., Perry, B. L., DeVerna, M. R., Yang, K-C, Flammini, A., Menczer, F. and Bryden, J., (2022), 'Online Misinformation is linked to Early Covid-19 Vaccine Hesitancy and Refusal', *Scientific Reports*, 12(5966). https://doi.org/10.1038/s41598-022-10070-w

Romer, D. and Jamieson, K. H., (2020), 'Conspiracy Theories as Barriers to Controlling the Spread of Covid-19 in the U.S.', *Social Science and Medicine*, 263: 113356.

Roozenbeek, J., Schneider, C. R., Dryhurst, S., Kerr, J., Freeman, A. L. J., Recchia, G., van der Bles, A. M., van der Linden, S., (2020), 'Susceptibility to Misinformation about Covid-19 Around the World', *Royal Society Open Science*, 7: 201199. https://doi.org/10.1098/rsos.201199

Russo, F. and Williamson, J., (2007), 'Interpreting Causality in the Health Sciences, *International Studies in Philosophy of Science,* 21(2): 157-170.

Salmon, W. C., (1998), *Causality and Explanation*, Oxford: Oxford University Press.

Shan, Y., (2022), 'Philosophical Foundations of Mixed Methods Research', *Philosophy Compass,* 17 (1). DOI: 10.1111/phc3.12804

Shan, Y. and Williamson, J., (2021), 'Applying Evidential Pluralism to the Social Sciences', *European Journal for the Philosophy of Science,* 11(4): 96

Shan, Y. and Williamson, J., (2023), *Evidential Pluralism in the Social Sciences, Routledge*

Sheeran, P., (2002), 'Intention-Behaviour Relations: A Conceptual and Empirical Review', *European Review of Social Psychology*, 12(1).

Southwell, B. G., Niederdeppe, J., Cappella, J. N., Gaysynsky, A., Kelley, D. E., Oh, A., Peterson, E. B., Chou, W-Y. S., (2019), Misinformation as a Misunderstood Challenge to Public Health', *American Journal of Preventive Medicine*, 57(2): 282-285.

Talia, B., (2022), 'Algorithms and Misinformation: The Constitutional Implications of Regulating Microtargeting', Fordham Intellectual Property, Media and Entertainment Law Journal, 32(4): 1107.

Talic, S., Shah, S., Wild, H., Gasevic, D., Maharaj, A., Ademi, Z., Li, X., Xu, W., Mesa-Eguiagaray, I., Rostron, J., Theodoratou, E., Zhang, X., Motee, A., Liew, D. and Ilic, D., (2021),

'Effectiveness of Public Health Measures in Reducing the Incidence of Covid-19, SARS-CoV-2 Transmission, and Covid-19 Mortality: Systematic Review and Meta-Analysis', *BMJ,* 375.

The Lancet Editorial, (2020), 'The Covid-19 Infodemic', *The Lancet Infectious Diseases*, 20(8): 875. Available at: The COVID-19 infodemic - The Lancet Infectious Diseases

Trofimov, A and Williamson, J., (forthcoming) 'Applying Evidential Pluralism to Evidence-Based Law: EBL+'.

Wardle, C. and Derakhshan, H., (2017), 'Information disorder: Toward an Interdisciplinary Framework for Research and Policy making', Council of Europe Report DGI(2017)09*, Council of Europe*. Available at: 168076277c (coe.int)

Weintraub, E. L., (2019) 'Opinion: Don't Abolish Political Ads on Social Media. Stop Microtargeting', *Washington Post*. Available at: Opinion | FEC Chair Ellen Weintraub: Don't ban political adds on social media. Ban microtargeting. - The Washington Post

Westerwick, A., Kleinnman, S. B. and Knobloch-Westerwick, S., (2013), 'Turn a Blind Eye if you Care: Impacts of Attitude Consistency, Importance and Credibility on Seeking Political Information and Implications for Attitudes', *Journal of Communication,* 63(3): 432-453.

What Works, (2018), 'The Rise of Experimental Government: Cross-Government Trial Advice Panel Update Report', Cabinet Office and Economic and Social Research Council. Available at: The Cross-Government Trial Advice Panel: Update Report - GOV.UK (www.gov.uk)

Woodhouse, J., (2020), 'Regulating Online Harms', Research Briefing, House of Commons Library. Available at: CBP-8743.pdf (parliament.uk)

World Economic Forum (2013*), Global Risks Eight Edition.* World Economic Forum (weforum.org)

World Health Organization, 'Infodemic'. Available at: Infodemic (who.int)

Biography: Alexandra Trofimov is a research associate in philosophy at the University of Kent, UK. Alexandra is currently working on a Leverhulme funded research project with Professor Jon Williamson that seeks to develop and defend a new approach to evidence-based law based on the principles of Evidential Pluralism. Alexandra has previously published in *Jurisprudence*.