Reading Between the Lies: A Classification Scheme of Types of Reply to Misinformation in Public Discussion Threads

George Buchanan

iSchool, University of Melbourne, Victoria, Australia, george.buchanan@unimelb.edu.au

Ryan Kelly

School of Computing and Information Systems, University of Melbourne, Victoria, Australia, ryan.kelly@unimelb.edu.au

Stephann Makri

City, University of London, London, U.K. stephann@city.ac.uk

Dana McKay

RMIT University, Melbourne, Victoria, Australia, danamckay@gmail.com

Online misinformation is a fiendish problem. Demonstrably false information propagates faster and more widely than truth and this has heralded a technological arms race. One possible mechanism for addressing misinformation is social: there is evidence seeing misinformation being challenged can 'inoculate' a reader against it. To date, no research has examined how discussions sparked by misinformation play out; What are the different ways in which people reply to posts containing misinformation? How does the discussion flow in each case? Are there differences between platforms? We address these questions through an inductive qualitative analysis of discussion threads on three public discussion platforms (Twitter, YouTube and two news sites) and on three topics (COVID, Brexit and climate change). We present a classification scheme of types of replies to misinformation, and show that replies show different patterns between platforms. Knowing how people reply to posts that contain misinformation enriches our knowledge of 'human misinformation interaction,' and provides an understanding of how socio-technical factors in platform design can reduce the risk of misinformation spreading.

CCS CONCEPTS • Information systems~Users and interactive retrieval • Human-centered computing~Human-Computer Interaction (HCI)

Additional Keywords and Phrases: Misinformation, information sharing, echo chambers

ACM Reference Format:

First Author's Name, Initials, and Last Name, Second Author's Name, Initials, and Last Name, and Third Author's Name, Initials, and Last Name. 2018. The Title of the Paper: ACM Conference Proceedings Manuscript Submission Template: This is the subtitle of the paper, this document both explains and embodies the submission format for authors using Word. In Woodstock '18: ACM Symposium on Neural Gaze Detection, June 03–05, 2018, Woodstock, NY. ACM, New York, NY, USA, 10 pages. NOTE: This block will be automatically generated when manuscripts are processed after acceptance.

1 Introduction

The spread of misinformation is one of the biggest challenges facing information interaction. Misinformation has been blamed for everything from the spread of COVID-19 to the outcome of elections [27; 52], and is the focus of much research: how it is created and by whom, what can be done to minimise its spread, and how to prevent people from believing it [14; 52; 70].

Types of misinformation, the actors involved in creating and spreading it, their motives, and the types of misinformation they spread have been thoroughly investigated [70]. Researchers have also examined how misinformation might be counteracted by machine learning and other automated approaches [42; 59]. However,

to effectively combat misinformation, we must address it from multiple perspectives [59; 60]. A particularly promising perspective is a social one; there is evidence seeing misinformation being challenged can 'inoculate' a reader against it [13]. However, the role people play in affirming or rejecting misinformation remains important and relatively under-researched area of 'human misinformation interaction' [60].

Other than who shares misinformation [21; 53], and that some people at least shake their heads and walk away [57], we know little about how people respond to misinformation in comment threads. As 50% of all our information needs are met by other people [54], the role discussion plays in responding to misinformation (e.g. in affirming or challenging it) is a surprising omission from the literature, particularly given that challenges to misinformation play a role in whether or not it is believed [13; 45].

We address this research gap by examining and categorising in-thread replies to comments containing misinformation on 3 public discussion platforms: Twitter, YouTube, and the comments sections of 'The Guardian' and 'Daily Mail' news sites. We examined multiple platforms to determine whether the nature and interface of each platform affected behaviour. We also examined 3 controversial topics on each platform: COVID-19, climate change, and Brexit, to ensure the types of replies to misinformation we identified were not topic-specific.

We first discuss prior work on user interactions with misinformation. We then explain and justify our research method. Next, we present a classification scheme of types of reply to posts containing misinformation. This is followed by a discussion of implications for information interaction research, and the design of future systems. Finally, we address limitations and future work.ACM's new manuscript submission template aims to provide consistent styles for use across ACM publications, and incorporates accessibility and metadata-extraction functionality necessary for future Digital Library endeavors. Numerous ACM and SIG-specific templates have been examined, and their unique features incorporated into this single new template. If you are new to publishing with ACM, this document is a valuable guide to the process of preparing your work for publication. If you have published with ACM before, this document provides insight and instruction into the current process for preparing` your manuscript.

2 Background

This section first discusses previous work on classifying and understanding discussion threads. Next we examine literature that defines misinformation, then addresses the persistent rise of misinformation, the role of people in its spread and existing approaches to combat it.

2.1 Information Interactions in Discussion Threads

We are far from the first researchers to attempt to classify responses in discussion threads. This research takes two forms: discussions of automated classification of discussion threads or individual posts, and qualitative analysis of posts to try to understand the informational content.

Automated discussion thread classifiers have been studied in both educational settings and social media [20; 33; 34]. These classifiers have been used to try to identify features of threads or posts such as aggression, trolling, announcements or question asking. Identifying these features has many practical uses for moderation, but these classifiers are a research tool in and of themselves. Manually classifying threads is labour intensive [33], and once key features of threads or posts have been identified automated classification can support qualitative analysis as in [68]. To identify these key features, though, requires just that labour intesnive initial manual coding; in this paper we offer an initial coding for misinformation discussion threads.

Turning to qualitative analyses, academic social networking sites, YouTube and news media sites have all been the subject of classification work. Sometimes, as with [47], analyses address sentiment (such as compliments), other analyses (e.g. [34; 68]) have focused on the informational content of posts or threads. The reflection on interactions between posters on these threads is often limited, Madden for example classifies them into agreement and disagreement [34]. Morrison considers the interactions between posters more deeply, but not systematically [41]. The idea that threads may in fact be misinformation is introduced in [41]. In that paper

Morrison notes that there is no way of verifying the personal anecdotes he describes, and that other material posted under news stories could equally be false. We examine discussion threads from a different perspective—not how do they inform, but how do they misinform, and what do people do when they read misinformation in these threads?

2.2 Defining Misinformation

Concisely defined, 'misinformation' is 'false or inaccurate information that is deliberately created and is intentionally or unintentionally propagated' [67]. Under this umbrella sits a range of types of misinformation, the definitions for which may be colloquial and overlapping. Two terms often seen in the literature are 'fake news' and disinformation. The distinction typically drawn between mis- and disinformation is that disinformation is spread deliberately, whereas misinformation is propagated without intent to mislead. However, determining intent to mislead is difficult, if not impossible [70]. Hence 'misinformation' is also used as an umbrella term to describe all demonstrably false information, regardless of the creator or sharer's intentions [67]. We use 'misinformation' in line with this definition.

An overlapping umbrella term is 'fake news', which refers to publicly incorrect or inaccurate information, particularly on social media [19; 60]. However, the types of information that fall under this term are contested. Scholars often include parody and satire [61]. While the public agree that biased information and propaganda are 'fake news', satire and parody are not seen the same way [43]. It is notable that, context free, satire can appear to be 'real news' to those not aware of the source (e.g. US satirical newspaper *The Onion*) [70]. A comprehensive literature survey concludes satire, bias, propaganda, clickbait, conspiracy theories, fabricated information, hoax and rumours can all be considered 'fake news' [70]. While some of these types of information are predominantly intended to mislead (e.g. propaganda, hoax, fabricated information) others (e.g. clickbait or rumours) may be either mis- or dis-information depending on the, often unascertainable, intent of the creator. This makes it difficult (if not impossible) to separate mis- from dis-information in practice. While the motivations of those who spread misinformation is an interesting research topic, it is not the focus of this paper, which is how people reply to misinformation, whatever its form.

2.3 The Rise and Rise of Misinformation

Widespread public misinformation is not new; propaganda campaigns during WWII gave rise to research on thought manipulation through misinformation [46].

Information scholars' focus on misinformation is relatively recent and has emerged from circumstances that appear to make its spread more likely. First, much news consumption now occurs via social media [18; 19], rather than directly from authoritative sources with obligations to maximise journalistic integrity [55]. While authoritative sources provide a de facto veracity check (and consumers recognise this [19]), the time required to verify every piece of information on social media is prohibitive [52].

Second, the visual distinction between fabricated content, news, and advertising is blurred on social media feeds, where the interface is controlled by the platform, rather than news organisation. Third, changing news media business models have meant the production of credible news content has become expensive and difficult to fund [2; 4]. In turn, news sites have been presenting advertising alongside news and moving towards clickbait-style links to articles, making it more difficult for news consumers to distinguish news from ads.

Fourth, misinformation is often designed to appeal to our emotions [70]. Some news consumers say they prefer emotional content to factual content, and so emotionally charged 'fake news' is more interesting than unbiased 'real news' [4; 36]. Incidentally this has led to some discussion in Digital Journalism about the ethics of using emotion to promote messages of social benefit [64], though this is not widespread practice.

Finally, technology drives misinformation. It is easier than ever to manipulate images [61], and there is a chance algorithmic bias may be driving spread of misinformation by promoting what is popular [4]. Similarly, technology democratises information production, but minimises the distinction between professionally

researched and edited information, and manipulated information or even lay opinion [70]. While we now better understand the rise of misinformation, what has not yet been considered is the ways replies to misinformation amplify, question or challenge it.

2.4 Spreading Misinformation

Misinformation is spread by two types of actor: bots (programmed machines supposed to fool humans into believing they are real), and people [70]. Of the people spreading misinformation, some do so deliberately because they are activists or trying to sow discord (this latter group colloquially known as trolls), or because they are paid to do so. Others spread it unintentionally, either because they are passionate about its message, or they are misguided and think it is interesting [70]. Still others know they are spreading misinformation, but think their reasons are relatively harmless, e.g. expressing opinions or starting conversations [10].

One early study found misinformation is not shared as regularly as the most ominous discussions would have us believe, but is most likely to be shared by politically conservative and older social media users [21]. More recent work found misinformation is shared because it is considered credible, and that credulity was associated with religious belief and conservative political ideology [58]. It is not just the old, right-wing and credulous who share misinformation though: recent work has demonstrated subscribers to a fact-checking newsletter share misinformation at rates similar to the rest of the population [53]. Still more concerning is that false information on Twitter spreads more widely than real information [63], which is why combating misinformation is both so difficult and important.

What these studies tell us overall is that social media users are more likely to share misinformation than truth, and that younger and politically liberal users are less likely than others to spread it. While we know who is spreading misinformation and why, we do not know how people are replying to it when they see it. This paper addresses that research question.

2.5 Approaches to Combatting Misinformation

Many approaches have been proposed to combat misinformation, including technical, educational and legal [52].

Algorithmic approaches fall into two categories: identifying known misinformation within a network, and neutralising it, or detecting it algorithmically. Algorithmic detection is useful but not sufficient to combat misinformation: new misinformation is constantly being created. Algorithmic predictors, on the other hand, must be trained on ground truth data [67]. This makes it very difficult to detect some kinds of misinformation (e.g. bias), because they are epistemologically difficult even for humans to classify [51]. If humans cannot classify misinformation, algorithms risk long term ineffectiveness or entrenched human bias [15]. The challenges of automated approaches to combating misinformation are noted in [59] and hybrid models proposed.

Another approach to mitigating misinformation is digital and information literacy training, so people recognise misinformation when they see it [14]. While information professionals have much to offer here, there are challenges with an information literacy approach [53; 60]. Most people are truth-biased: they believe what they read, and that they are better at discerning truth than they actually are [32; 52]. Also, once a belief is entrenched, it is surprisingly difficult to dislodge [32].

A third approach is legislation, requiring social media organisations to monitor and reduce misinformation [2; 23; 25]. This has been widely posited as a social good, but the question of how organisations might police this remains; they face all the difficulties we have already raised.

A final approach, most relevant to our study is social: rather than preventing people from seeing or sharing misinformation, it relies on those who do see it attacking its credibility. The value of this approach stems from the finding that others' posts may mitigate against both the spread of and belief in misinformation. Comments that challenge misinformation decrease trust in it [6; 13]. Detailed information about why something is incorrect results in an even more effective challenge [9]; providing an alternative narrative also works [32]. Given that vastly more people read posts and comments than write them in an online forum [44], educating lurkers so that

they do not become believers or sharers could be a very powerful strategy. The effect of challenges to misinformation is not universal, though: they can result in a backlash [45], as found by psychological research [32].

While we know others' posts may mitigate the spread of misinformation, we do yet know what the different types of replies to posts containing mining actually are and how discussions sparked by misinformation play out within threads. Our paper addresses this research gap.

3 Method

We examined replies to posts that contained demonstrably false misinformation in public discussion threads. We considered misinformation as content we could demonstrate to be wholly or partly factually incorrect or inaccurate. While videos or articles containing misinformation are rare, (platforms often remove them) comments on content that themselves contain misinformation, are surprisingly abundant. Hence, we focused on this more commonplace form of misinformation.

3.1 Data Collection and Fact Checking

We examined replies on public discussion platforms to posts that contained misinformation on 3 controversial, non-overlapping topics (COVID-19, climate change and Brexit) and on 3 platforms that facilitate open discussion on public interest issues: Twitter, YouTube, and two news sites The Guardian (theguardian.com) and Daily Mail (dailymail.co.uk). We chose *controversial* topics because they frequently result in discussions that contain misinformation. We chose multiple *non-overlapping* topics to increase the likely generalisability of our findings.

The platforms represent both social media and mainstream websites where information sharing and discussion occurs, around a variety of content types (e.g. images, news, video). The newspapers have different political stances (the Guardian centre-left, Daily Mail right-wing). While Facebook is a common platform for hosting misinformation [21; 36], it was excluded from this research as most content, particularly misinformation [56] is private, and analysing private posts breaches its terms of service.

News sites provide some degree of quality control of the core content on which users can comment, and the comments on both sites we examined are moderated. Twitter and YouTube host user-generated content and provide minimal moderation (only in response to users reporting comments). While users may sign up to and post anonymously on all these platforms, and all posts are publicly visible, those on Twitter can be searched within the platform, while those on YouTube and online news sites cannot. Although Twitter posts have a maximum length of 280 characters, comments on the other platforms are unrestricted in length, though often as short as Tweets. Finally, users see content specifically from those they follow in Twitter, while YouTube shows a mixture of recommended and searched content. The news sites studied do not personalise news content.

We focused on COVID-19, climate change and Brexit because media coverage suggests misinformation on these topics is rife [5; 16; 52] this focus allowed us to discover posts containing misinformation relatively readily. In contrast to debate on gun control and capital punishment, these issues do not divide cleanly across partisan lines, and polarised views are found across the political spectrum [1; 5; 26].

On each platform, we conducted internal searches using the terms "COVID-19", "climate change" and "Brexit" in September and October 2020. We examined the results list for each platform and selected the first results that met our inclusion criteria: that is the first post contained demonstrably false information and had at least one reply (as we were investigating replies to misinformation). For the Guardian and the Daily Mail, this was 5 posts each per topic, resulting in a set of 30 posts (i.e., 2 news sites, 3 topics, 5 posts each). For Twitter and YouTube, this was 10 posts each per topic resulting in a further 60 posts (2 platforms, 3 topics, 10 posts each). This involved examining 40-100 threads per search, as many did not contain misinformation. This resulted in a dataset of 90 posts in total containing misinformation and 847 replies to those posts. On Twitter, 'posts' were Tweets containing misinformation. On YouTube, and the news sites, 'posts' were comments on content (videos or news stories) The searches were conducted in September and October 2020.

To establish whether an initial post contained misinformation, we sought confirmation of unlikely-looking statements via Google to see if evidence could be found to support each claim, supplemented by reference to fact-checking websites such as Snopes. A second researcher (one of the authors) verified any posts been flagged as misinformation by the first researcher and only those they agreed on were included in the analysis.

The second researcher took care to verify candidate misinformation independently, to avoid potential agreement bias. As will be seen, most misinformation that passed these tests was egregiously erroneous. We allowed a maximum of 15 minutes' verification time to reflect the limited time taken by readers in practice [52]. Misinformation that requires sophisticated verification tactics to identify is likely to be accepted as truth [32], and thus to provoke different patterns of replies. In practice, no checks exceeded this time limit. While it is possible some instances of misinformation were missed, this does not impact the validity of our findings as it did not prevent us from taking an equal sample across types of topic and discussion platform.

3.2 Data Analysis

For each misinformation thread, all content that was clearly spam or robot-generated (e.g. where there were multiple posts containing the same textual content from the same account) was excluded from analysis. Images that were text-rich were transcribed, but profile images were discarded.

The individual replies were then coded inductively as to their nature, e.g. mocking, affirmation. This coding continued until data saturation was achieved, i.e. no new reply types were being identified. This general inductive approach is commonly used in information interaction research to identify types of information behaviour (e.g. see [40]). To ensure rigour, if a new type of reply was identified, the entire dataset was re-reviewed for all reply types. Thread-level patterns of replies were identified by following chains of content within threads, considering the initial codes and poster identities. We also examined all replies made by each individual poster to gain a detailed understanding of individual posts in the context of their other replies. We tested agreement for coding of reply types between the first and a second coder (with 97.4% inter-rater agreement).

Previous research has tested the relative frequency of observed patterns in information artefacts [17; 37]. The appropriate test for this study is log-linear analysis, that supports three-dimensional analysis of frequency data (the more common chi-squared test only permits two) across platform, topic and reply type.

3.3 Ethical Considerations

Our results are based on public data. It is good ethical practice to paraphrase posts when reporting such data because people do not expect their posts to be used out of context [3], and our research in particular may seem to posters to cast aspersions on their behaviours. While it is impossible to ensure complete anonymity when reporting public posts, as with enough search effort it may be possible to identify posters, we used obscuration to conceal users' identities. We had to strike a careful balance, however, as obscuring the data too much can confuse or distort meaning. This approach was approved by our ethics board

4 Results

We first describe the types of individual *reply* to misinformation found in our data. These types are found *within* an individual post (on Twitter) or comment (on YouTube, The Guardian and Daily Mail). Although distinct, there are semantic similarities between some types (e.g. mockery and insult). We then discuss the *thread-level* common patterns of discourse found in replies to a post with misinformation. Posters' identities are given in encoded form: 'OP' is the poster of the original comment containing misinformation; 'RPn' is given for the nth poster giving a reply (e.g. RP1 is the poster of first reply). We did not encounter any case where a poster or a responder changed their opinion. We tested for platform differences and report them where significant.

4.1 Types of Reply to Misinformation Posts

We identified 24 types of reply to posts containing misinformation. These fell into three groups: 1) **introducing misinformation and bias**, where replies added their own misinformation, conspiracy theories and racism; 2) **responding to points raised**, a large group of 16 separate codes such as affirmation, rebuttal, insult, quotation, and flooding and 3) **drawing on evidence**, which comprised personal experience, anecdotes, external sources and expertise and asking users to 'Google' a topic, or certain information to check its veracity. We now briefly define each type, referring to examples.

4.1.1 Introducing Further Misinformation and Bias.

Misinformation: As well as the original posts, 114 replies contained identifiable misinformation. They could introduce misinformation either in support of, or against, the original post. A reply on Twitter Brexit Thread 8 claimed (British) lawyers get to choose cases to represent, but lawyers are allocated cases, and they are required to accept and represent any client, regardless of their personal view of the client or case.

Conspiracy theory: these were often found in both the original post with misinformation, and in replies that approved of the original post (166 replies). Conspiracy theories were occasionally found in replies that took the opposite view to the original poster (6 replies): e.g. when replying to a climate-change denying original post, a critic claimed that climate-change was intentionally caused by the Koch Brothers and their supporters.

Racism: 36 posts included racist, xenophobic or anti-Semitic content. This was particularly the case with Brexit (29 posts), but also co-occurred with conspiracy theories in general. An example from Twitter Climate Change Thread 4 said: "Attenborough has been a shill for George Soros for decades. He almost certainly knows that the CC hoax is all about Agenda 2030 and reducing world population." (Agenda 2030 refers to UN work on sustainable development, often referred to by anti-Semitic conspiracy theorists as a plot against Western Democracies [66]).

4.1.2 Responding to Points Raised.

Posts responding to misinformation could **affirm** it (128 replies), either by repeating or simply agreeing; **rebut** it (198 replies), by an explicit disagreement with it; **question** it (100 replies), by asking for more information, clarification or interrogating its completeness or accuracy e.g. *"I'd need you to expand on that for me to believe you?"*; engage in **mockery** (102 replies, only 10 of which were newspaper comments), *"only a fool would…"* or **satire** (44 replies); *"so many epidemiologists here!"*; or simply **insult** e.g. *"you're a prime idiot*" (51 replies, 33 on YouTube alone).

Quotation: This was where authors directly or indirectly referred to comments previously posted in the thread, by themselves or others. In 42 posts the poster quoted their own original post (e.g. "as I said at the beginning…"), and 15 quoted another author's post (e.g. "as you already said…").

There were also a number of reply types that diverted the discussion: there were 18 cases of **whataboutery**, where a poster raised a point that distracted from one point by raising another that was plausibly, but not actually, relevant. An example of this is a poster asking *"what about the twenty thousand that die each day?"* diverting from a debate about COVID restrictions. There were 69 **asides** that presented a different issue (e.g. one poster saying they needed something to eat). A relatively sophisticated type was the **side-step** (48 posts). These began with a point raised by the poster, but then would invite further comment on a specific issue (e.g. *"I agree that it can be hard to get a job, but is there just one problem?"*), which could be an attempt to tease out the original poster's deeper reasoning on factors influencing governmental lockdown decisions. Four posts **misrepresented** an earlier post, by distorting context, changing the words, or making claims not in the original (a localised form of misinformation) to serve the interests of the current poster: e.g., in YouTube Climate Change Thread 10, RP1 responded to the OP's claim that all science is politically motivated *"Is the periodic table a hoax too? What will the radical right criticise about atomic weights next year?"*, the OP misrepresented RP1's post as confirming his claim, stating *"RP1 I'm happy to see that you confirm that climate science is fraudulent."*

There was a group of reply types that did not directly express a particular view on the topic. A rare reply type found in 6 supporting and contrarian posts, was a statement that matters were **complex**. While sometimes this may have been a form of side-step, in others it appeared more to serve a full-stop in a debate. A related type was **fatalism** (17 replies), agreeing with a previous post, but suggesting that little could be done about an issue raised in it. A third form was **speculation** (39 replies), in which posters ruminated about possible future states of affairs. In Twitter COVID Thread 8, a poster speculated about President Trump's COVID-19 treatment *"he's likely taking the latest experiment to show the world there's a cure and COVID's not going to kill you"*.

Not just facts were questioned: some posters agreed with the data presented in a post, but questioned the poster's **reasoning** (39 replies), e.g. *"the EU can't be both manifestly inept and efficiently manipulative and cunning—so which is it?"*, or, in the context of climate change, an OP used absolute data for China and the USA, and stated this data showed the pollution per person; after multiple clarifications, a respondent asked *"if you understand it, why did your initial post state the absolute opposite?"*.

Flooding*:* This was a relatively rare (10 replies) type that included large amounts of diverse information, potentially serving to confuse or overwhelm users. Flooding was found in the both the original post containing misinformation, and in supportive and critical replies, apparently to contradict a particular point with which the poster strongly disagreed: e.g. in YouTube Brexit Thread 10, RP6 posted a reply with 3 URLs and 13 paragraphs.

4.1.3 Drawing on Evidence.

Posters intermittently drew on a variety of sources to support their point of view. Two common examples were **personal experiences** (14 replies) and **anecdotes** (6 replies). Personal experience was invoked when a poster claimed a personal encounter with an issue (e.g. *"my mother had COVID earlier this year"*), while anecdotes were either not explicitly reported personally, or were clearly from third parties, e.g. *"more people in that town speak Romanian than English"*, or *"I'm told that the hospitals are all empty"*. Clearly, we cannot verify that an experience really occurred, or people are who they claim to be.

References to external sources or expertise were marginally more common. One type was an direct appeal to a named authority (19 replies), e.g. *"David Attenborough said that.."*; or an indirect appeal without specifics (16 replies), e.g. *"any scientist will tell you that"*, or *"my doctor told me"*. Particular material could be referenced by a URL or clear label, such as the title and author or a book. We divided these references into **true references** i.e. the material was relatively reliable and said what was claimed of it (10 replies); and **false references** where it was unreliable, inaccurate or did not say what was claimed (9 replies). False references are a potential form of misinformation and it is notable that misuses of references were almost as common as accurate uses. Eight replies simply directed others to **Google** a particular piece of information, if they wanted to check it for themselves. The most common form of evidence was **testable facts**: material which, while not referenced, could in principle be tested for accuracy, e.g. the number of people who had influenza in 2018. There were 77 replies including testable facts.

4.2 Thread-Level Patterns of Replies

Beyond individual replies to posts containing misinformation, we identified 7 common patterns that emerged across entire discussion threads. The patterns were:

Drive-by: the original poster of misinformation makes a comment, but never returns to the discussion.

Echo chamber: all replies agree with the original poster, and none disagree with or correct the original misinformation.

Piling on: one poster, often the OP, receives a series of replies from multiple accounts critiquing the post or point of view.

Ding-dong: two posters take opposite positions and exchange a long sequence of posts, making and rebutting claims and counterclaims, often with little movement.

Broken record: a poster repeatedly posts the same or highly similar content in reply to multiple counter-points.

Non-answer: a poster is asked a question, but repeatedly explicitly or implicitly avoids answering it.

Counter-attack: a poster is challenged on a point and replies with insults, mockery, or suggestions that the questioner(s) are complicit in, or fooled by, a conspiracy.

Non-answer and counter-attack threads were relatively rare, and are thus not reported here. Where original posters did return to their threads, broken record, non-answer and counter-attacks were more common than clarifications or engagement. The other thread types are reported here, with reasons why the OP was misinformation provided at the end of each example.

4.2.1 Drive-By Threads.

In these threads, the original poster (OP) never replied to their original post. An example is Twitter Climate Thread 8, where the OP posted *"two minutes is too little time to show the gap between natural cycles of the climate and the Socialist, Anti-Business Hoax they call 'man-made climate change'".* This resulted in 31 replies, with no further comment from the OP.

4.2.2 Echo-Chamber Threads.

An example echo-chamber comes from Twitter Brexit Thread 3. Although this thread has some asides (by RP2 and RP3), the flow of the conversation echoes and even amplifies the OP's view that all lawyers who represent migrants in asylum cases can be considered activists. Therefore, this is a case of an *echo-chamber,* where the original post containing misinformation is never corrected, despite 7 users replying. Instead, RP4 and RP5 agree strongly and vociferously with the OP, affirming the OP's view and others amplify the original post by further insulting lawyers who represent asylum requests, e.g. by calling them *"globalist traitors"* (RP1), *"well-dressed thieves"* (RP1) and *"rascals"* (RP7) and by extending the criticism introduced by the OP to institutions as well as individual lawyers (RP4, RP6). RP7 also amplifies the OP rather than merely affirming it, by arguing that lawyers who take on asylum cases drain UK taxpayers:

OP: "Supreme court's chair says barristers who represent migrants' requests for asylum aren't activists, but simply doing their job. The Home Secretary is right: they're all activists. Selfish, money-grabbing ones too. (newspaper URL)." **[misinformation]**

RP1: "Globalist traitors one and all: they hate the UK, hate Brexit" **[insult, affirmation]**

RP2&3: **[both present asides on unity]**

RP4: "Scots Parliament, Welsh Assembly, Supreme Court: All set up by Blair. If I ran the UK we'd abolish the lot." **[affirmation]**

RP5: "Agreed – they are so far out of touch." **[affirmation]**

RP6: "The 'supreme' court is just another bunch of socialist activists. Add 'em to the list, they all betray us." **[affirmation]**

RP7: "Well-dressed thieves [insult]; English taxpayers have nothing left, these scum [insult] take everything". **[affirmation]**

30 later replies RP8-30 all **affirm** the original post.

Why is the OP misinformation? It ignores the basic mechanism of assigning lawyers in the UK legal system: lawyers seldom choose their cases.

4.2.3 Piling-On Threads.

Here, the OP receives no, or almost no support. Replies are consistently negative and critical. For example, in the newspaper Climate Change thread #3, responding to an article about the energy needs of transport in the future, the OP claims wind power requires fossil fuel backup, while solar power is 'worthless.' RP1 is simply dismissive, saying these claims *"don't match the science"* and RP2 piles on by providing an example of useful solar power. RP3 and RP4 also pile onto the discussion. RP3 argues against the OP's view that *"wind power needs coal power plants…for fall-back capacity,"* mocking the OP's claims as *"whisps in the wind".* RP3 states it is actually the other way around: electric vehicle turbines have been used as a backup to the National Grid's non-renewable electricity supply. RP4 feeds off RP3's post on electric vehicles as a source of renewable energy, inviting the OP to *"Google"*

smart grids and electric vehicles. Cumulatively, these replies serve to undermine the credibility of the misinformation.

OP: "Wind power needs coal power plants running all day for fall-back capacity. They don't solve the problem of coal, gas and carbon dioxide; deforestation caused by wood-powered stations are ridiculously unsustainable, even after a century. Solar power is worthless and doesn't work in our latitudes - they break down within a decade". **[misinformation]**

RP1: "Your claims don't fit the science. One bit." **[rebuttal]**

RP2: [car manufacturer] built their UK engine factory not far from where I live – they're now starting to make electric engines there. The top of this (massive) factory is covered with over 20,000 solar panels, producing c. 6MW. They often sell the energy they don't need to the national grid **[personal experience]**. You seem to be thinking about how solar was in the 1970s!" **[mockery]**.

RP3: "Let me show how your ideas are whisps in the wind **[mockery]**. Turbines on electric vehicles are a demonstrated backup for the grid. **[rebuttal, testable fact]**. I don't actually like the concept, because it assumes we'll happily give our own power over to the grid, unless we all have to lease the batteries from it. Anyhow, that's not the only solution!" **[complex]**.

RP4: "You should Google 'smart grids': large-scale ownership of electric vehicles is an integral part of the network." **[Google]**

Why is the OP misinformation? The need for permanently operational fossil fuel plants has been shown to be unnecessary by several innovations, including some of the those reported by the respondents. Wood-based electric power production is still little-used, but much improved in efficiency since its inception.

4.2.4 Ding-Dong Threads.

These featured extended, often heated exchanges between the same pair of posters, sometimes with a gradual shift in the discussion, but often with little movement on either side: discussions often ended in a stalemate. This example is from YouTube COVID Thread 8, where RP1 and the OP engage in a long exchange, focused on the logic of the original (incorrect) claim that the UK public *"said no to"* the March 2020 COVID-19 lockdown. This exchange demonstrates very little movement as the OP re-phrases their original claim to counter RP1's criticism, but RP1 replies by questioning the OP's reasoning. Furthermore, both the OP and RP1 engage in mockery and insult:

OP: "The nation said no to any lockdown!" **[misinformation]**

RP1: "You represent yourself, not the public: I'm one of them. **[rebuttal, personal experience]** You've problems!" **[insult]**

OP: "I've problems with what is being done. Many don't fall for the hoax. **[conspiracy theory]**. I meant to say most of the nation **[misinformation]** …; not morons like you." **[insult]**

RP1: "You changed to 'many' you started saying 'the nation'; **[questioning reasoning]**. I highlight it and the only thing you can do is to respond with nasty, puerile name-calling; if honesty hurts, put up with it mate!" **[mockery]**.

OP: "Brighten up!" **[mockery]** followed by an **aside** then exchange continues for another 5 posts, returning to the original point. Posters RP2 and RP3 then reply in support of the OP.

Why is the OP misinformation? There was no public vote on UK movement restrictions. Opinion polls showed most (88%, IPSOS/MORI) favoured a lockdown when the video was recorded.

4.2.5 Broken Record Threads.

In broken record threads the OP repeatedly makes the same argument they have already made), without further evidence, despite counter-arguments from other posters. In this example from YouTube Brexit Thread 8, the original video is from the BBC and discusses the UK's position in Brexit negotiations, after having left the EU on 1st Jan. 2020. The OP repeats their view the UK has not really left the EU, having only left *"on paper,"* while, RP2 repeatedly replies that they have as *"if we did leave on paper, we have then left"*. Both sides play their own broken record in a deadlocked discussion that, unlike ding-dongs, does not involve making and replying to evolving points and counter-points.

OP: "It's on paper: it's not really true. We haven't really left the European Union; unfair chances? You mean freedom to do business!" **[misinformation]**
RP1: "You signed up on paper – one you want to hide **[rebuttal, mockery].** Odd how your leaders want to violate international treaties. Very reliable, sure!" **[satire, mockery]**.
OP: "We voted to leave 4 years ago. We've not really left. Our decision is being blocked. Who cares about BS paper and lying politicians. We voted out, and not to trade any longer." **[misinformation, testable fact]**.
RP2: "If we did leave on paper, we have then left. The UK isn't in the EU anymore. If a couple sign divorce papers, they're divorced." **[rebuttal, questioning reasoning]**.
OP: "We still can't control our borders, commerce, farms. It's obvious it's all a lie. We voted out four years back. We didn't vote to leave only in name." **[misinformation, conspiracy theory]**.
RP2: "We are out. Full stop. I don't care what folks voted for or against…what they feel they can do. We ain't in the EU any more, so we're out! That's just a simple fact." **[rebuttal, testable fact].**
RP2: "You voted out of the EU; you left; no vote said 'leave and no discussions, no trade agreements'". **[rebuttal, testable fact]**.
OP: "RP2 I know you don't care what I voted for. You're just another politician. **[insult, mockery]**. If we get a deal we won't be leaving. **[misinformation]**. Your post is so EU. You don't care about democracy **[insult]**. Whatever the vote, the EU does what it wants" **[conspiracy theory]**.
RP2: "We left already, that's the truth, move on!" **[rebuttal]**
OP: "We still can't control our borders, commerce, farms. It's obvious it's all a lie. We voted out over four years back. We didn't vote to leave only in name. I don't know where you're from but I believe in democracy. You don't." **[**possible **racism, insult]**.
RP2: "You didn't vote to leave and no have no talks; You voted out of the EU; you left the EU". [Note that both this and the OP's last post are almost identical to previous posts; the exchange continues for a further 25 posts].

Why is the OP misinformation? The UK was in an ambiguous position; outside the EU, but in an agreed one-year transition period. The UK had left the EU and was free to engage in activities it was not allowed to do as an EU member, e.g. making independent trade agreements with states outside the EU. The original BBC video explicitly stated this.

4.3 Thread Types and Platforms

The relative frequency of the different types of thread is shown in Table 1 below. Note that more than one interaction style may be seen in each thread. For example, where the original poster does a 'drive by', the posters responding to it may end up in a 'ding-dong'. Thus, the column totals do not add up to 30.

Table 1: Distribution of Thread Types by Platform (All counts are out of 30)

Platform	Twitter	YouTube	Newspaper
Drive-by	24	14	21
Echo chamber	16	5	4
Pile on	1	12	16
Ding-dong	5	9	5
Counter-attack	2	6	3
Broken record	3	13	6
Non-answer	1	1	2

We first tested to check if the distribution was non-random. The overall log-linear analysis test for interactions is strongly significant (G^2=291.5, df=32, p<0.0001), indicating that there is some underlying pattern. There is a statistically significant difference between platforms, and platforms are the dominant factor of difference (G^2=251, df=14, p<0.0001). Comparing the platforms reveals the detailed differences: Twitter has a high number

of echo-chamber threads, and few 'pile-ons' (χ^2=14,73, df=2, p=0.0006), while YouTube has few 'drive-by's (χ^2=11.40, df=2, p=0.0033), but many broken-record threads (χ2=8.44, df=2, p=0.0146).

Turning to reply types, we also found differences. Due to space limitations, we only present the most marked differences here. For example, misinformation was significantly more prevalent in YouTube replies (72 of 293) than newspapers (35 of 170) or Twitter (72 of 384 replies) (χ^2=22.7, df=2, p<0.0001).

In terms of response styles, Twitter had many affirmations, just over 25% of replies (97 of 384), versus less than 10% on YouTube (15 from 293) and newspaper comments (16 of 170) (χ^2=56.0; df=2, p<0.0001). In contrast, insults were rare on both Twitter (14 replies), and newspapers (4) but higher on YouTube (33, 11.6%) (χ^2=23.36, df=2, p<0.0001). Mockery, on the other hand was common on both Twitter and YouTube (53 and 39 cases respectively), but rare in newspapers replies (10) (χ^2=7.89, df=2, p=0.0193).

5 Discussion

In this section we first hypothesize reasons for the different interaction patterns on different platforms, then compare threads to search and other forms of information seeking. Finally, we discuss the design implications of our work.

5.1 Different Platforms, Different Interactions

The different interactions by platform are clear. Twitter threads lack discussion: they are predominantly 'drive by' or 'echo chamber' threads, so even where there is dissent it is from a single post. At the post level there is a disproportionate amount of affirmation. This conversational makeup reflects differences between Twitter and the other platforms: threads on Twitter are most likely to be viewed by followers with whom one has some form of relationship. These relationship ties mean posters are plausibly less likely to see things they disagree with [7; 12], and less likely to react strongly negatively to it if they do. [57]. In contrast to Twitter, both YouTube and newspapers had a high proportion of pile on threads, though YouTube threads were more likely to include some form of discussion (ding dong or broken record) than newspapers, and included more insults and misinformation. Insults on YouTube have been seen before [34] specifically noted insults, and advice that may actually also be insult (e.g. 'you should wear less makeup') YouTube is colloquially known for having generally contentious, negative discussion threads, and our findings support that assessment [31]. In contrast newspaper threads (which are moderated) are more likely to offer extensive correction of misinformation in the comments, potentially mitigating against its spread [13; 32]. Posts on newspapers are less likely to include conspiracy theory, mockery or racism, probably due to moderation. This more 'civil discourse' [69] is more conducive to the types of discussion that may dislodge false beliefs in both posters and lurkers [13; 32; 45]. The key message here, though, is not *why* these different platforms have such varied response patterns, but *that* they do, thus demonstrating that the culture and interactive properties of a platform can and do influence social responses to misinformation.

5.2 Threads as Information Seeking

In this section we contribute to the discussion of discussion threads in information interaction. We extend previous work looking at the informational content of posts (e.g. [41; 47; 68]) instead drawing analogies between interactions at the thread level and other types of information interaction. By comparing threads to search and other forms of information seeking we focus specifically on how they might support people in questioning misinformation.

For most thread-level patterns of replies to misinformation, there is an analogous behaviour in the information interaction literature. The drive-by post is akin to 'grab-and-go' information acquisition, where the poster has a single target- to say or find something specific, but not engage [8; 38]. Piling on is the equivalent of non-diverse search results or recommendations, where replies are very similar and thus not very interesting [11; 29; 30]. While interactive to some degree (the search may evolve based on the information found, just as replies with

further misinformation when piling on may build on and feed off each other), piling on is not highly iterative. The broken record pattern is a more extreme example of the same: it is like users retyping almost identical queries and expecting a different result [28].

Perhaps more usefully for combating misinformation, ding dongs are like exploratory search, which is a nuanced and highly-interactive dialogue between system and searcher where the information interaction evolves in response to the information returned by the system, sometimes in exciting new directions [65]. Learning is a key outcome of exploratory search, and has potential to be with ding dongs too.

It is interesting that, like exploratory search, the ding dong discussion is the one most likely to promote learning: people who see diverse perspectives are those most likely to change their views [6; 9; 32; 40], which could potentially involve them reconsidering the credibility or accuracy of information they had held as true. While we did not see any changes of view manifest publicly in the data, viewpoint change is a complex process that is unlikely to occur in a single thread [40]. As noted above, Twitter is particularly poor at promoting this kind of interaction, whereas newspaper sites are good at it. Previous research suggests that if thread-level replies to misinformation could be a respectful discussion conducted with the spirit of openness and understanding, rather than a 'ding dong,' the effects would be twofold. First, it would likely invite more diverse reply posts in turn; many people do not want to argue and only want to engage if there is a good chance of a reasonable discussion [57]. Second, a discussion with nuance and respect is more likely than a ding dong to promote open-minded deliberation [5; 6; 40] which, in turn, could potentially disrupt the spread of misinformation.

These analogies highlight the potential for participants in these discussion forums to treat the thread as a search interface, testing their thoughts and ideas This analogy further frames the discussion as a document for a silent third party, or lurker, who may use what is said in the thread to discern the veracity of information presented. Given that lurkers represent 90%+ of online groups [44], this potential use is too large to be ignored. This framing highlights the potential value of the social response to misinformation: participants in discussion forums debunking it, not for the benefit of those with whom they are discussing, but for the benefit of the silent watchers to reduce their trust and increase their reasoning and accuracy focus [13; 48; 49].

5.3 Design Implications

Ideally information interfaces will support the spread of truth, rather than falsehood. Our research suggests efficacy in two possible ways of doing this: flagging and removing potential misinformation (the algorithmic approach), and supporting the challenge of misinformation.

The patterns we have identified are potentially amenable to systematizing, offering one potential means for at least algorithmically flagging discussion threads that may be misinformation [42; 52], an activity that has been done for other post and thread types in the past [20; 62]. While misinformation threads would still likely need human checking [59], this would provide one means of identifying what should be checked. It remains an open question for information science, whether once misinformation is identified, removing it entirely, or flagging it as suspicious but allowing it to stand alongside replies is a more effective means of combatting misinformation. On one hand, misinformation that has been removed cannot be shared; on the other seeing misinformation debunked regularly supports readers in developing a critical and questioning mindset, thus improving recognition of misinformation generally [13; 48].

Differences in reply patterns across platforms provide lessons for the design of public discussion threads specifically, social media platforms, and information environments more broadly. The first is diversity begets diversity: public discussion platforms aimed at a diverse audience (e.g. news sites) and presenting diverse information without little personalisation are those with the most diverse and wide-ranging discussions. This diversity offers the alternative narratives, reasoning pathways, and challenges to misinformation that can promote reflection among readers and those who may join in the discussion alike [13; 32; 48; 49], turning a potentially negative interaction (where misinformation is encountered) into a positive one (misinformation is encountered alongside a counterargument, promoting learning).

It is a key feature of our study that, with the possible exception of the Twitter data, all the original posts in our dataset were encountered by those who replied: there is no way of searching the comments, highlighting the importance of the information encounter in misinformation interaction The decisions made by the designers of the platforms we studied manipulate which information is likely to be encountered—comments with recent replies or lots of 'upvotes' are likely to appear near the top.. The role of information encounters in promoting diversity and new perspectives has been discussed before [35; 39; 50], though not in the context of misinformation. It would be easy to reach for forcing diversity as a protection against misinformation, [24], this will not work. For an information encounter to be experienced as a positive rather than negative 'distraction,' the encounterer's mind must be open to that information [35]. Seeing one's beliefs disagreed with alone is not enough to promote re-examination of those beliefs [13; 45]: the information needs to be relatable enough to engage with and the information interface needs to inspire a sense of curiosity.

Our challenge as information interaction specialists is twofold. The first issue is old: generating better support for digital information encountering, which despite increased research is a poorly supported type of information acquisition [39]. The second is new: the CHIIR community is ideally placed to identify how readers of news and social media feeds might be provided with information that is just diverse enough to inform, but not so diverse as to alienate. The ethics of such nudges toward diverse consumption may be debatable, but promoting reflection (rather than using algorithms to persuade) is widely accepted [22]. This approach is also the most likely to be effective in combatting an entrenched, but false, viewpoint [32].

6 Limitations and Future Work

We recognise the relatively small sample (by quantitative standards) used in this primarily qualitative study is only a starting point for further quantitative research. Analysis at scale may reveal further misinformation reply types and enable stronger claims about their general prevalence or importance. Automation could assist with large-scale analysis; the broken record and ding-dong patterns, for example, might be automatically identified based on textual repetitions. This could make quantitative analysis at scale more practical.

While peoples' replies to posts containing misinformation were a possible indication of whether they were likely to propagate the misinformation (e.g. by endorsing it), we did not examine their behaviour beyond their replies; future work might investigate whether users exposed to certain types of reply to misinformation are more or less likely to spread it.

We examined 3 topics across 3 platforms, yielding a variety of post-level reply types and thread-level patterns. A wider variety of platforms and topics could lead to further types of reply being identified and reveal nuances within particular types. Differences between our chosen platforms and more private venues, such as closed Facebook or WhatsApp groups, could prove enlightening.

Furthermore, we only examined misinformation: it is plausible contrarian replies to *correct* information might reveal different patterns, but further qualitative work is needed. While we saw no users publicly change their view about misinformation, understanding the specific reply types and/or interactions around those critical moments may yield additional valuable insights into how to mitigate the threat of misinformation.

7 Conclusions

To address the fiendish challenge of misinformation, it is necessary to understand its 'human nature.' Understanding misinformation from a social perspective, through a deeper understanding of 'human-misinformation interaction' on platforms that facilitate information sharing and discussion is a particularly promising approach. We present the first classification of types of reply to misinformation on public discussion platforms, addressing both individual posts, and entire threads. This in turn provides a foundation for reasoning about online debates: e.g. about how best to encourage readers to post counters to misinformation. It also

provides an empirically-grounded theoretical lens for informing future research on social interactions with misinformation.

We have also identified several implications for information interaction. The first is a new potential way of automatically identifying misinformation-containing discussion threads. The second is the role of discussion interfaces in promoting challenges to misinformation. The third is the importance of information encountering in addressing misinformation, and the role of discussion platforms in nudging toward or away from certain encounters.

Understanding how people reply to misinformation can enrich our understanding of 'human misinformation interaction' and, in turn, reduce its power and spread.

References

<bib id="bib1"><number>[1]</number>Ashcroft, L., 2016. How the United Kingdom voted on Thursday… and why, Lord Ashcroft Polls. https://lordashcroftpolls.com/2016/06/how-the-united-kingdom-voted-and-why/</bib>
<bib id="bib2"><number>[2]</number>Australian Competition and Consumer Commission, 2018. Digital Platforms Inquiry: Preliminary Report. https://www.accc.gov.au/focus-areas/inquiries-finalised/digital-platforms-inquiry-0.</bib>
<bib id="bib3"><number>[3]</number>Ayers, J.W., Caputi, T.L., Nebeker, C., and Dredze, M., 2018. Don't quote me: reverse identification of research participants in social media studies. Digital Medicine 1, 1 (2018/08/02), 30. DOI= http://dx.doi.org/10.1038/s41746-018-0036-2.</bib>
<bib id="bib4"><number>[4]</number>Bakir, V. and McStay, A., 2018. Fake News and The Economy of Emotions. Digital Journalism 6, 2 (2018/02/07), 154-175. DOI= http://dx.doi.org/10.1080/21670811.2017.1345645.</bib>
<bib id="bib5"><number>[5]</number>Benegal, S.D. and Scruggs, L.A., 2018. Correcting misinformation about climate change: the impact of partisanship in an experimental setting. Climactic Change 148, 1 (2018/05/01), 61-80. DOI= http://dx.doi.org/10.1007/s10584-018-2192-4.</bib>
<bib id="bib6"><number>[6]</number>Bode, L. and Vraga, E.K., 2018. See Something, Say Something: Correction of Global Health Misinformation on Social Media. Health Communication 33, 9 (2018/09/02), 1131-1140. DOI= http://dx.doi.org/10.1080/10410236.2017.1331312.</bib>
<bib id="bib7"><number>[7]</number>Bruns, A., 2019. Filter bubble. Internet Policy Review 8, 4. DOI= http://dx.doi.org/10.14763/2019.4.1426.</bib>
<bib id="bib8"><number>[8]</number>Buchanan, G. and McKay, D., 2011. In the Bookshop: Examining Popular Search Strategies. In Proc. JCDL 11 (Ottawa, Canada), ACM, 269-278. DOI= http://dx.doi.org/10.1145/1998076.1998127.</bib>
<bib id="bib9"><number>[9]</number>Chan, M.-p.S., Jones, C.R., Hall Jamieson, K., and Albarracín, D., 2017. Debunking: A Meta-Analysis of the Psychological Efficacy of Messages Countering Misinformation. Psyc. Sci 28, 11 (2017/11/01), 1531-1546. DOI= http://dx.doi.org/10.1177/0956797617714579.</bib>
<bib id="bib10"><number>[10]</number>Chen, X., Sin, S.-C.J., Theng, Y.-L., and Lee, C.S., 2015. Why Do Social Media Users Share Misinformation? In Proc. JCDL 15 (Knoxville, Tennessee, USA), Association for Computing Machinery, 111–114. DOI= http://dx.doi.org/10.1145/2756406.2756941.</bib>
<bib id="bib11"><number>[11]</number>Clarke, C.L., Kolla, M., Cormack, G.V., Vechtomova, O., Ashkan, A., Büttcher, S., and MacKinnon, I., 2008. Novelty and diversity in information retrieval evaluation. In Proc. SIGIR 08, ACM, 659-666. DOI= http://dx.doi.org/10.1145/1390334.1390446.</bib>
<bib id="bib12"><number>[12]</number>Colleoni, E., Rozza, A., and Arvidsson, A., 2014. Echo chamber or public sphere? Predicting political orientation and measuring political homophily in Twitter using big data. J Comm 64, 2, 317-332.</bib>
<bib id="bib13"><number>[13]</number>Colliander, J., 2019. "This is fake news": Investigating the role of conformity to other users' views when commenting on and spreading disinformation in social media. Comp. Hum. Behav. 97(2019/08/01/), 202-215. DOI= http://dx.doi.org/10.1016/j.chb.2019.03.032.</bib>
<bib id="bib14"><number>[14]</number>Connaway, L.S., Julien, H., Seadle, M., and Kasprak, A., 2017. Digital literacy in the era of fake news: Key roles for information professionals. ASIST Proceedings 54, 1, 554-555. DOI= http://dx.doi.org/10.1002/pra2.2017.14505401070.</bib>
<bib id="bib15"><number>[15]</number>Cowgill, B., Dell'Acqua, F., Deng, S., Hsu, D., Verma, N., and Chaintreau, A., 2020. Biased Programmers? Or Biased Data? A Field Experiment in Operationalizing AI Ethics. In Proc. EC20 (Virtual Event, Hungary), Association for Computing Machinery, 679–681. DOI= http://dx.doi.org/10.1145/3391403.3399545.</bib>
<bib id="bib16"><number>[16]</number>Cuan-Baltazar, J.Y., Muñoz-Perez, M.J., Robledo-Vega, C., Pérez-Zepeda, M.F., and Soto-Vega, E., 2020. Misinformation of COVID-19 on the Internet: Infodemiology Study. JMIR 6, 2 (2020/4/9), e18444. DOI= http://dx.doi.org/10.2196/18444.</bib>
<bib id="bib17"><number>[17]</number>Fakis, A., Hilliam, R., Stoneley, H., and Townend, M., 2013. Quantitative Analysis of Qualitative Information From Interviews: A Systematic Literature Review. J Mixed Methods Res 8, 2 (2014/04/01), 139-161. DOI= http://dx.doi.org/10.1177/1558689813495111.</bib>
<bib id="bib18"><number>[18]</number>Fletcher, R. and Nielsen, R.K., 2017. Are people incidentally exposed to news on social media? A comparative analysis. New Media & Society 20, 7 (2018/07/01), 2450-2468. DOI= http://dx.doi.org/10.1177/1461444817724170.</bib>
<bib id="bib19"><number>[19]</number>Flintham, M., Karner, C., Bachour, K., Creswick, H., Gupta, N., and Moran, S., 2018. Falling for Fake News: Investigating the Consumption of News via Social Media. In Proc. CHI 18 (Montreal QC, Canada), ACM, New York, NY, 1-10. DOI= http://dx.doi.org/10.1145/3173574.3173950.</bib>
<bib id="bib20"><number>[20]</number>Fornacciari, P., Mordonini, M., Poggi, A., Sani, L., and Tomaiuolo, M., 2018. A holistic system for troll detection on Twitter. Comput. Hum. Behav. 89, 258-268. DOI= http://dx.doi.org/10.1016/j.chb.2018.08.008.</bib>
<bib id="bib21"><number>[21]</number>Guess, A., Nagler, J., and Tucker, J., 2019. Less than you think: Prevalence and predictors of fake news dissemination on Facebook. Science Advances 5, 1, eaau4586. DOI= http://dx.doi.org/10.1126/sciadv.aau4586.</bib>
<bib id="bib22"><number>[22]</number>Hansen, P.G. and Jespersen, A.M., 2013. Nudge and the manipulation of choice: A framework for the responsible use of the nudge approach to behaviour change in public policy. Eur J Risk Regulation 4, 1, 3-28.</bib>

<bib id="bib23"><number>[23]</number>Helberger, N., 2011. Diversity by Design. J. Inf Policy 1, 441-469. DOI= http://dx.doi.org/10.5325/jinfopoli.1.2011.0441.</bib>
<bib id="bib24"><number>[24]</number>Helberger, N., Karppinen, K., and D'Acunto, L., 2018. Exposure diversity as a design principle for recommender systems. Inf. Comm & Soc. 21, 2 (2018/02/01), 191-207. DOI= http://dx.doi.org/10.1080/1369118X.2016.1271900.</bib>
<bib id="bib25"><number>[25]</number>Helberger, N., Kleinen-von Könisglöw, K., and van der Noll, R., 2015. Regulating the new information intermediaries as gatekeepers of information diversity. Information and Media 17, 6, 50-71.</bib>
<bib id="bib26"><number>[26]</number>Ibbetson, C., 2020. Brits support new lockdown rules, but many think they don't go far enough, YouGov.</bib>
<bib id="bib27"><number>[27]</number>Islam, M.S., Sarkar, T., Khan, S.H., Mostofa Kamal, A.-H., Hasan, S.M.M., Kabir, A., Yeasmin, D., Islam, M.A., Amin Chowdhury, K.I., Anwar, K.S., Chughtai, A.A., and Seale, H., 2020. COVID-19–Related Infodemic and Its Impact on Public Health: A Global Social Media Analysis. Am. J Tropical Medicine & Hygiene 103, 4, 1621-1629. DOI= http://dx.doi.org/10.4269/ajtmh.20-0812.</bib>
<bib id="bib28"><number>[28]</number>Jansen, B.J. and Spink, A., 2006. How are we searching the World Wide Web? A comparison of nine search engine transaction logs. IP&M 42, 1, 248-263. DOI= http://dx.doi.org/10.1016/j.ipm.2004.10.007.</bib>
<bib id="bib29"><number>[29]</number>Kaminskas, M. and Bridge, D., 2016. Diversity, Serendipity, Novelty, and Coverage: A Survey and Empirical Analysis of Beyond-Accuracy Objectives in Recommender Systems. ToIS 7, 1, 1-42. DOI= http://dx.doi.org/10.1145/2926720.</bib>
<bib id="bib30"><number>[30]</number>Knijnenburg, B.P., Willemsen, M.C., Gantner, Z., Soncu, H., and Newell, C., 2012. Explaining the user experience of recommender systems. User Modeling and User-Adapted Interaction 22, 4-5, 441-504.</bib>
<bib id="bib31"><number>[31]</number>Lange, P.G., 2021. Conceptualizing communities of truth on YouTube. Explorations in Media Ecology 20, 1, 33-54.</bib>
<bib id="bib32"><number>[32]</number>Lewandowsky, S., Ecker, U.K.H., Seifert, C.M., Schwarz, N., and Cook, J., 2012. Misinformation and Its Correction: Continued Influence and Successful Debiasing. Psyc Sci in Public Interest 13, 3 (2012/12/01), 106-131. DOI= http://dx.doi.org/10.1177/1529100612451018.</bib>
<bib id="bib33"><number>[33]</number>Lin, F.-R., Hsieh, L.-S., and Chuang, F.-T., 2009. Discovering genres of online discussion threads via text mining. Comput & Ed. 52, 2 (2009/02/01/), 481-495. DOI= http://dx.doi.org/10.1016/j.compedu.2008.10.005.</bib>
<bib id="bib34"><number>[34]</number>Madden, A., Ruthven, I., and McMenemy, D., 2013. A classification scheme for content analyses of YouTube video comments. J Doc 69, 5, 693-714. DOI= http://dx.doi.org/10.1108/JD-06-2012-0078.</bib>
<bib id="bib35"><number>[35]</number>Makri, S., Blandford, A., Woods, M., Sharples, S., and Maxwell, D., 2014. "Making my own luck": Serendipity strategies and how to support them in digital information environments. JASIST 65, 11, 2179-2194. DOI= http://dx.doi.org/10.1002/asi.23200.</bib>
<bib id="bib36"><number>[36]</number>Marchi, R., 2012. With Facebook, Blogs, and Fake News, Teens Reject Journalistic "Objectivity". J Comm Inquiry 36, 3 (2012/07/01), 246-262. DOI= http://dx.doi.org/10.1177/0196859912458700.</bib>
<bib id="bib37"><number>[37]</number>Marshall, C.C., 1998. Toward an ecology of hypertext annotation. In Proc. HT 98 (Pittsburgh, Pennsylvania, USA), Association for Computing Machinery, 40–49. DOI= http://dx.doi.org/10.1145/276627.276632.</bib>
<bib id="bib38"><number>[38]</number>McKay, D., Chang, S., Smith, W., and Buchanan, G., 2019. The Things We Talk About When We Talk About Browsing: An Empirical Typology of Library Browsing Behavior. JASIST 70, 12, 1383-1394. DOI= http://dx.doi.org/10.1002/asi.24200.</bib>
<bib id="bib39"><number>[39]</number>McKay, D., Makri, S., Chang, S., and Buchanan, G., 2020. On Birthing Dancing Stars: The Need for Bounded Chaos in Information Interaction. In Proc. CHIIR 2020 (Vancouver BC, Canada), Association for Computing Machinery, 292–302. DOI= http://dx.doi.org/10.1145/3343413.3377983.</bib>
<bib id="bib40"><number>[40]</number>Mckay, D., Makri, S., Gutierrez-Lopez, M., MacFarlane, A., Missaoui, S., Porlezza, C., and Cooper, G., 2020. We are the Change that we Seek: Information Interactions During a Change of Viewpoint. In Proc. CHIIR 20 (Vancouver BC, Canada), Association for Computing Machinery, 173–182. DOI= http://dx.doi.org/10.1145/3343413.3377975.</bib>
<bib id="bib41"><number>[41]</number>Morrison, J., 2017. Finishing the "Unfinished" Story. Digital Journalism 5, 2 (2017/02/07), 213-232. DOI= http://dx.doi.org/10.1080/21670811.2016.1165129.</bib>
<bib id="bib42"><number>[42]</number>Nguyen, N.P., Yan, G., Thai, M.T., and Eidenbenz, S., 2012. Containment of misinformation spread in online social networks. In Proc. WebSci '12 (Evanston, Illinois), Association for Computing Machinery, 213–222. DOI= http://dx.doi.org/10.1145/2380718.2380746.</bib>
<bib id="bib43"><number>[43]</number>Nielsen, R.K. and Graves, L., 2017. "News you don't believe": Audience perspectives on fake news. https://ora.ox.ac.uk/objects/uuid:6eff4d14-bc72-404d-b78a-4c2573459ab8.</bib>
<bib id="bib44"><number>[44]</number>Nonnecke, B. and Preece, J., 2000. Lurker demographics: counting the silent. In Proc. CHI 00 (The Hague, The Netherlands), Association for Computing Machinery, 73–80. DOI= http://dx.doi.org/10.1145/332040.332409.</bib>
<bib id="bib45"><number>[45]</number>Nyhan, B. and Reifler, J., 2010. When Corrections Fail: The Persistence of Political Misperceptions. Political Behaviour 32, 2 (2010/06/01), 303-330. DOI= http://dx.doi.org/10.1007/s11109-010-9112-2.</bib>
<bib id="bib46"><number>[46]</number>O'Donnell, V. and Jowett, G.S., 1992. Chapter 4. In Propaganda and Persuasion Sage, 122-154.</bib>
<bib id="bib47"><number>[47]</number>Ostermaier-Grabow, A. and Linek, S.B., 2019. Communication and Self-Presentation Behavior on Academic Social Networking Sites: An Exploratory Case Study on Profiles and Discussion Threads on ResearchGate. JASIST 70, 10 (2019/10/01), 1153-1164. DOI= http://dx.doi.org/10.1002/asi.24186.</bib>
<bib id="bib48"><number>[48]</number>Pennycook, G., Epstein, Z., Mosleh, M., Arechar, A.A., Eckles, D., and Rand, D.G., 2021. Shifting attention to accuracy can reduce misinformation online. Nature 592, 7855 (2021/04/01), 590-595. DOI= http://dx.doi.org/10.1038/s41586-021-03344-2.</bib>
<bib id="bib49"><number>[49]</number>Pennycook, G. and Rand, D.G., 2019. Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. Cognition 188(2019/07/01/), 39-50. DOI= http://dx.doi.org/10.1016/j.cognition.2018.06.011.</bib>
<bib id="bib50"><number>[50]</number>Reviglio, U., 2017. Serendipity by design? How to turn from diversity exposure to diversity experience to face filter bubbles in social media. In International Conference on Internet Science Springer, 281-300.</bib>
<bib id="bib51"><number>[51]</number>Robertson, R.E., Jiang, S., Joseph, K., Friedland, L., Lazer, D., and Wilson, C., 2018. Auditing Partisan Audience Bias within Google Search. In Proc. CSCW 18 (Austin, TX), New York NY, 1-22. DOI= http://dx.doi.org/10.1145/3274417.</bib>

<bib id="bib52"><number>[52]</number>Rubin Victoria, L., 2019. Disinformation and misinformation triangle: A conceptual model for “fake news” epidemic, causal factors and interventions. J Doc 75, 5, 1013-1034. DOI= http://dx.doi.org/10.1108/JD-12-2018-0209.</bib>
<bib id="bib53"><number>[53]</number>Saling, L.L., Mallal, D., Scholer, F., Skelton, R., and Spina, D., 2021. No one is immune to misinformation: An investigation of misinformation sharing by subscribers to a fact-checking newsletter. PLOS ONE 16, 8, e0255702. DOI= http://dx.doi.org/10.1371/journal.pone.0255702.</bib>
<bib id="bib54"><number>[54]</number>Savolainen, R., 1995. Everyday life information seeking: Approaching information seeking in the context of “way of life”. L&ISR 17, 3 (1995/06/01/), 259-294. DOI= http://dx.doi.org/10.1016/0740-8188(95)90048-9.</bib>
<bib id="bib55"><number>[55]</number>Schudson, M., 2001. The objectivity norm in American journalism*. Journalism 2, 2 (2001/08/01), 149-170. DOI= http://dx.doi.org/10.1177/146488490100200201.</bib>
<bib id="bib56"><number>[56]</number>Scott, M., 2020. Facebook’s private groups are abuzz with coronavirus fake news. In Politico Axel Springer, Germany.</bib>
<bib id="bib57"><number>[57]</number>Seargeant, P. and Tagg, C., 2019. Social media and the future of open debate: A user-oriented approach to Facebook’s filter bubble conundrum. Doiscourse Context and Media 27(2019/03/01/), 41-48. DOI= http://dx.doi.org/10.1016/j.dcm.2018.03.005.</bib>
<bib id="bib58"><number>[58]</number>Stefanone, M.A., Vollmer, M., and Covert, J.M., 2019. In News We Trust? Examining Credibility and Sharing Behaviors of Fake News. In Proc. SMS19 (Toronto, ON, Canada), Association for Computing Machinery, 136–147. DOI= http://dx.doi.org/10.1145/3328529.3328554.</bib>
<bib id="bib59"><number>[59]</number>Strickland, E., 2018. AI-human partnerships tackle "fake news": Machine learning can get you only so far-then human judgment is required. IEEE Spectrum 55, 9, 12-13. DOI= http://dx.doi.org/10.1109/MSPEC.2018.8449036.</bib>
<bib id="bib60"><number>[60]</number>Sullivan, M.C., 2018. Why librarians can’t fight fake news. JLIS 51, 4 (2019/12/01), 1146-1156. DOI= http://dx.doi.org/10.1177/0961000618764258.</bib>
<bib id="bib61"><number>[61]</number>Tandoc Jr, E.C., Lim, Z.W., and Ling, R., 2018. Defining “fake news” A typology of scholarly definitions. Digital Journalism 6, 2, 137-153.</bib>
<bib id="bib62"><number>[62]</number>Ventirozos, F.K., Varlamis, I., and Tsatsaronis, G., 2018. Detecting Aggressive Behavior in Discussion Threads Using Text Mining. In Proc. CICLing 18 (Mexico City), Springer International Publishing, Berlin, Germany, 420-431. DOI= http://dx.doi.org/10.1007/978-3-319-77116-8_31.</bib>
<bib id="bib63"><number>[63]</number>Vosoughi, S., Roy, D., and Aral, S., 2018. The spread of true and false news online. Science 359, 6380, 1146-1151. DOI= http://dx.doi.org/10.1126/science.aap9559.</bib>
<bib id="bib64"><number>[64]</number>Wahl-Jorgensen, K., 2016. Emotion and journalism. In The SAGE Handbook of Digital Journalism SAGE, 128-143.</bib>
<bib id="bib65"><number>[65]</number>White, R.W. and Roth, R.A., 2009. Exploratory search: Beyond the query-response paradigm. Synthesis lectures on information concepts, retrieval, and services 1, 1, 1-98.</bib>
<bib id="bib66"><number>[66]</number>Wilson, J., 2016. Agenda 21 is conspiracy theory. But don't dismiss Malcolm Roberts as a harmless kook, The Guardian. Guardian Media Ltd., London, UK, https://www.theguardian.com/commentisfree/2016/sep/14/agenda-21-is-conspiracy-theory-but-dont-dismiss-malcolm-roberts-as-a-harmless-kook.</bib>
<bib id="bib67"><number>[67]</number>Wu, L., Morstatter, F., Carley, K.M., and Liu, H., 2019. Misinformation in Social Media: Definition, Manipulation, and Detection. SIGKDD Explor. Newsl. 21, 2, 80–90. DOI= http://dx.doi.org/10.1145/3373464.3373475.</bib>
<bib id="bib68"><number>[68]</number>Yarmand, M., Yoon, D., Dodson, S., Roll, I., and Fels, S.S., 2019. "Can you believe [1:21]?!": Content and Time-Based Reference Patterns in Video Comments. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems Association for Computing Machinery, Paper 489. DOI= http://dx.doi.org/10.1145/3290605.3300719.</bib>
<bib id="bib69"><number>[69]</number>Yom-Tov, E., Dumais, S., and Guo, Q., 2013. Promoting Civil Discourse Through Search Engine Diversity. Soc Sci Comp Rev 32, 2 (2014/04/01), 145-154. DOI= http://dx.doi.org/10.1177/0894439313506838.</bib>
<bib id="bib70"><number>[70]</number>Zannettou, S., Sirivianos, M., Blackburn, J., and Kourtellis, N., 2019. The Web of False Information: Rumors, Fake News, Hoaxes, Clickbait, and Various Other Shenanigans. J. Data and Inf. Quality 11, 3, Article 10. DOI= http://dx.doi.org/10.1145/3309699.</bib>