"The headline was so wild that I had to check": A mixed-methods exploratory analysis of women’s health misinformation on social media

Lisa Mekioussa Malki
HCl-E MSc Final Project Report 2022
UCL Interaction Centre, University College London
Supervisors: Aneesha Singh and Dilisha Patel

Abstract
The circulation of health misinformation on social media is a growing socio-technical problem which has received considerable attention in recent years. However, despite evidence that misinformation affects women and men differently, the element of gender has largely been disregarded in existing HCI literature. This dissertation seeks to ameliorate the lack of scholarly work focusing on gendered health misinformation by conducting a multi-platform investigation into the topics, sources, and formats of women’s health misinformation currently circulating online, and its impact on female users.

First, a content analysis of 191 women’s health-related social media posts flagged as misinformation by fact-checking organisations is conducted. Then, a diary study is carried out with 19 female participants to investigate women’s encounters with misinformation on social media using Dervin’s sensemaking methodology as an analytical frame. Study 1 finds that most of the officially fact-checked women’s health misinformation is related to reproductive health, whereas Study 2 suggests that women are most likely to encounter and be negatively impacted by weight loss misinformation. Study 2 also reveals a breadth of strategies used by women to identify and make sense of health misinformation, finding that inter-subjective sensemaking is facilitated by socio-technical affordances, and features such as comment sections and engagement metrics. These results have the potential to inspire design interventions which support women in navigating health misinformation online.

CCS CONCEPTS
• Human-centered computing → Empirical studies in HCI.

Keywords
Women’s health, misinformation, social media, sensemaking

MSC. CONTRIBUTION TYPE
Empirical

MSC HCI-E FINAL PROJECT REPORT
Project report submitted in part fulfilment of the requirements for the degree of Master of Science (Human-Computer Interaction with Ergonomics) in the Faculty of Brain Sciences, University College London, 2022.

NOTE BY THE UNIVERSITY
This project report is submitted as an examination paper. No responsibility can be held by London University for the accuracy or completeness of the material therein.

1. INTRODUCTION
The use of misinformation to promote political causes, drum up military support, and stoke anti-scientific agendas predates the digital era, but there is little doubt that social media provides new opportunities for false information to spread at scale [23]. In recent years, a large quantity of misinformation pertaining to the COVID-19 pandemic was seen to circulate online, resulting in skewed public perceptions towards governmental handling of the virus [36, 69] and non-compliance with public health measures such as mask-wearing and vaccination [27]. As such, social media has been dubbed ‘ground-zero’ in the fight against health misinformation. Though the COVID-19 ‘infodemic’ has dominated this research landscape in recent years, a range of women’s health issues have also seen a rise in inaccurate or misleading information online. For instance, misinformation surrounding abortion [52], body image and weight loss [20], HPV vaccination [11], and breast cancer [81] has been found on a range of platforms.

It is appropriate to view misinformation exposure through a gendered lens, as structural inequalities have been found to shape people’s everyday interactions with information [76]. For instance, exploratory research in rural India has found that women are sometimes coerced by male relatives into sharing COVID-19 misinformation against their will [79]. Elsewhere, an Australian survey found that women are more likely than men to use the Internet to seek out health information when they are experiencing “stigmatised conditions” such as sexual and mental health issues [63]. Much of the dedicated research into women’s health misinformation has focused on analysing samples of informative posts on social media [20, 52, 81] and investigates either a single platform or health topic in isolation. While valuable, these studies do not consult users, and therefore give little empirical insight into the subjective experience of women exposed to this content, its wider social impact, and how these impacts can be mitigated.

Furthermore, existing misinformation research is frequently conducted in artificial settings, and lacks a critical consideration of the user’s situational context [64]. This is problematic, as “situational triggers” and subtle contextual factors can influence user decisions to engage with, and share misinformation on social media [75]. One such contextual factor is the socio-technical environment in which misinformation is embedded. The affordances and network structures of different social media platforms can configure different behavioural outcomes for users, and have been found to render some platforms more ‘fertile grounds’ for misinformation propagation than others [75, 83]. Thus, exploratory, in-situ research may better capture the naturalistic contexts in which misinformation
is encountered, and uncover a fuller breadth of situational and technical nuances which shape the user’s experience [64].

In sum, this dissertation addresses three key gaps in the literature: the lack of realistic, in-situ insights into how users react to health misinformation on social media; the largely disregarded element of gender; and a view of social media platforms that frames them as an incidental host of misinformation, rather than an active factor which shapes and constrains its spread. To address the research gaps outlined above, two studies were designed. First, a content analysis was conducted to characterise the common types and formats of women’s health misinformation which has seen recent circulation on social media. Secondly, a diary study was carried out to investigate how women identify and react to online health misinformation in their everyday lives using Dervin’s sensemaking methodology [21] as a theoretical framework. Both studies represent novel undertakings, as no known studies to date have quantified women’s health misinformation across multiple platforms, or used an in-situ, qualitative method to investigate women’s experiences with health misinformation.

The structure of the dissertation is as follows: first, related literature will be reviewed, with a focus on previous attempts at quantifying health misinformation; gender dynamics; the impact of platform features and affordances on the proliferation of misinformation content; existing misinformation interventions; and Dervin’s sensemaking methodology. Then, the research questions, methods, and results of the two studies will be described. Finally, the results will be critically appraised in light of the existing literature, and areas for further study will be proposed.

2. LITERATURE REVIEW

2.1 Terminology

The terms ‘misinformation’ and ‘disinformation’ are separated according to criteria of veracity and intentionality [40], where disinformation refers to content which is intentionally designed to be misleading or false [70], and misinformation is used to either imply a lack of intent to do harm [34], or to discuss false information regardless of creator intent [28, 40]. The term misinformation is used throughout this study to remain neutral to the poster’s intentions, as this can be contentious and difficult to determine in a research context [67]. By extension, the term ‘women’s health misinformation’ is used to refer to misinformation that is either related to health issues which are “unique to, more prevalent in, or manifest differently in women” [44], or have been identified by women themselves as health priorities. The latter category goes beyond discrete diseases or conditions, and encapsulates the entire spectrum of health concerns perceived to be important by women, including structural issues such as gendered bias in healthcare provision [17].

2.2 Quantifying Health Misinformation

Existing research reveals a diversity of actors involved in the propagation of misinformation online, as well as a spectrum of veracity along which posts are situated. Different types of misinformation have been found to vary in prevalence and impact, with tacit ‘half-truths’ constituting 59% of a sample of 225 COVID-19 related posts flagged as misinformation by official fact-checkers [69]. The same study found that while the majority of misinformative posts were bottom-up, i.e., originated from members of the public, a sizable minority of posts were created or shared by celebrities, and attracted a disproportionate share of social media engagement [69]. Larger-scale studies support this finding—an automated review of 1500 Tweets officially flagged as misinformation found that verified and celebrity Twitter accounts often act as ‘super-spreaders’ of fabricated claims [67].

The use of officially fact-checked posts has clear advantages over other approaches to collecting samples of misinformation, despite the bottleneck it imposes. Some automated approaches use source reliability to identify misinformation [2, 15], but this is problematic, as misinformation has been known to originate from sources deemed trustworthy, or from lone individuals [67]. Other approaches include comparing the contents of social media posts to guidelines published by official health bodies such as the World Health Organisation (WHO) [45]. Despite bypassing the various selection biases of fact-checkers, this approach places the burden of judging whether a particular post is misinformation on the researcher: a time-consuming and potentially bias-prone process. Using the verdicts of fact-checkers therefore, is a more robust and scalable approach to sample collection.

2.3 Misinformation and Women’s Health

A growing body of literature draws attention to the unstable bodily autonomy experienced by women over time, and how this is reflected in long-standing medical discrimination [26]. Recently, young women suffering from endometriosis have been found to experience dismissal, delayed diagnosis, and a sense of not ‘feeling heard’ by medical staff [82], and similar patterns have emerged for women with polycystic ovary syndrome (PCOS) [37]. Disbelief of female medical complaints has resulted in a deep-rooted mistrust of medical authorities, underpinning many women’s decisions to adopt alternative or holistic forms of care which employ a rhetoric of female empowerment [41]. However, these so-called ‘new wellness’ movements can promote pseudoscientific health treatments, and have been identified by medical organisations as a risk to women’s reproductive and physiological health [6].

Gendered dynamics operate not only in the topics of misinformation people are exposed to [3], but also in how information is presented visually. One recent study reviewed the prevalence of weight loss misinformation on Pinterest—an image-based platform with an 82% female userbase [20], finding that of a sample of 234 pins relating to misinformative belly fat loss exercises, almost three quarters featured images of young, slim women in bathing suits [20]. Similarly, over a quarter of these images appeared to be filtered or visually enhanced. This illustrates that while weight loss misinformation may not exclusively target women, its framing, presentation, and impact can be gendered.

Lastly, gender has been found to impact how people respond to misinformation on a cognitive level, though existing research paints a mixed picture of how this manifests in practice. Some surveys suggest that women consume social media content more critically than men [3, 84], whereas other studies find that women are less skeptical of misinformation and more likely to share it [14, 80]. This
An affordance is broadly described as a possibility for action, arising in the "relational structure between an object or technology, and the user." [24]. Affordance theory can be used to explain how platform design imposes constraints on a user’s actions in light of their individual needs and motivating factors [75]. For instance, the affordances of replicability and scalability can facilitate the spread of misinformation by making it easy to share content in such a way that makes it difficult to distinguish the original from its copy [12]. The potential visibility of this content is large, so can be disseminated in ‘informational cascades’ and reach a wide audience before it is removed or corrected [75]. Other relevant affordances include anonymity, as users may interact with misinformation differently depending on whether their activity is visible to friends [75], and moderation, which can configure the ideological breadth of content shown to users [83].

Though not visible to the user, the underlying architectures of social media play a powerful role in configuring how users interact with information. Algorithms which recommend content to users frequently prioritise the likelihood of engagement over accuracy or diversity of opinion [83], and can perpetuate algorithmic gender bias and lack transparency in their outcomes [1]. Even search engines are only as neutral as their information retrieval systems allow: decisions on which results to show users often prioritise relevance and commercial optimisation over accuracy [35, 59]. Users generally lack a concrete understanding of how these ‘black-box’ systems work [71], and despite attempting to construct folk-theories to explain their underlying mechanics [22], experience feelings of uncertainty and distrust towards algorithmic mediation [64].

A well-discussed drawback of fact-checking is the ‘information deficit’ model, which assumes that belief in misinformation stems from a lack of scientific knowledge, and can be corrected by the provision of facts [68]. This disregards the interpersonal and attentionally demanding context of social media, where users frequently experience scattered attention, emotional burnout, and informational overload [34, 36]. To cope with the perceived cost of critically evaluating web content, people use peripheral cues and cognitive heuristics to make rapid, rule-based judgements of credibility in such a way that optimises information utility as a function of interaction cost [47]. For instance, users are more willing to accept misinformation corrections that are issued by friends [38] or consistent with their worldview [31]. Fact-checking therefore, is unlikely to solve the ‘infodemic’ on its own, as it neglects the interpersonal and affective factors underlying belief in misinformation [23].

More novel interventions leverage behavioural nudges at posting time to encourage more mindful sharing of content on social media. Facebook and Twitter have recently deployed popups which appear when a user attempts to share an article they have not read, to encourage media literacy and limit the spread of misinformative and inflammatory content [33]. Though there is little published empirical insight into the effectiveness of these interventions, other lightweight nudges which operate at the time of posting have previously been found to reduce the intent to share misinformation in large scale experiments [38].

Sensemaking refers to the cognitive and behavioural processes by which people come to understand a complex or ambiguous reality [48]. Though initially developed to analyse organisational structures and practices, sensemaking metatheory has since been used to analyse how people managed the information landscape of the COVID-19 pandemic [58, 59]. It has also seen direct application to misinformation, in studies which investigate the role played by platform affordances on the proliferation of conspiratorial content online [83].

Dervin’s sensemaking methodology (SMM) [21] consists of four key interacting elements: gaps, bridges, outcomes, and the situational context of the individual [57]. The framework’s explicit focus on contextual factors makes it an ideal deductive tool for understanding user cognition once exposed to misinformation, which as established, is influenced by the context of social media. Central to the methodology is the metaphysical concept of a gap, defined as a discontinuity or disruption to a person’s understanding that arises during unexpected and irregular situations, of which exposure to misinformation is an example [72]. Cognitive gaps mark a discrepancy between what is known, and what a person feels should be known, driving a pressing need to clarify, understand, and correctly ’frame’ ambiguous data. Gaps are ‘bridged’ using internal resources, or by engaging in information seeking strategies [59]. Bridges eventually produce an outcome, defined as a momentary clarity or understanding of a situation. This is not a static or stable state, as the sensemaking process continually sees people subjected to new situations and information requiring constant reassessment. [48]. This formulation is illustrated in Figure 1, with contexts, gaps, bridges, and outcomes all experienced as a user moves through time.
The key data collection method associated with Dervin’s SSM is the flexible micro-moment timeline interview [21, 57]. The core technique involves asking a participant to recount a specific situation step-by-step and using content-free, neutral questions to probe the gaps experienced at each stage in more depth. This method has been deployed in an elicitation interview context [57–59] as well as in diary studies as a framework for entries [21].

Figure 1: Core SSM triangle metaphor, reproduced from [57]

3. RESEARCH QUESTIONS

As outlined in the literature review, the sources, framing, and visual presentation of misinformation can have significant impacts on how it is received by users. Hence, the aim of the first study is to characterise the topical focus, sources, types, and media formats of a sample of social media posts relating to women’s health, which have been marked as misinformation by official fact-checkers. The study also analyses the extent to which different social media platforms have restricted the spread of posts in the sample, through fact-checking captions, post deletion, and account suspension.

RQ1: What are the key types, sources, formats, and claims of women’s health misinformation identified by fact-checkers on social media?

RQ2: To what extent have social media companies intervened in the spread of women’s health misinformation on their platforms?

The second study addresses the lack of in-situ research into how women make sense of health misinformation on social media. Dervin’s sensemaking framework is used to attend methodologically to how women frame and process misinformation, and the and socio-technical factors shaping their experience. Despite well-researched differences in behaviour and cognition across men and women [14, 80], meta-analyses have revealed a gender gap in HCI research. Women are underrepresented as research participants in both classic and recent studies [51], meaning that gendered stereotypes and perceptions of men as the default user may be sidelining women’s needs as active users of social media. This study, therefore, centres the experiences and perspectives of women to ensure their equal participation in the fight against health misinformation.

RQ3: What cognitive gaps, concerns, and uncertainties arise from women’s encounters with health misinformation on social media?

RQ4: Which sensemaking strategies are used by women to identify and process health misinformation, and what are their outcomes, for instance, in terms of decisions to share or report content?

RQ5: How do the features and affordances of different social media platforms influence the sensemaking processes described in RQs 3 and 4?

4. STUDY 1: CONTENT ANALYSIS

4.1 Methods

4.1.1. Data Collection

The Google Fact-check Explorer API was used to collect records of all indexed fact-check articles relevant to women’s health published between July 2021 and July 2022. A Python script was used to query the API on 31st July 2022 with a set of keywords informed by previous literature [17, 56, 86] and the US Office of Women’s Health’s list of health topics [50]. This source was chosen on account of its recency, and its comprehensive coverage of a range of health topics known to primarily affect or manifest differently in women, including reproductive, physical, and mental health issues, weight loss and dieting, autoimmune diseases, and particular cancers. The API returned a JSON response, containing an entry for each fact-check article. The variables extracted from each entry are summarised in Table 1 below.

Table 1: Summary of variables extracted from the JSON response.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Claim</td>
<td>A 100-200 character summary of the misinformation claim being made by the fact-checked post.</td>
</tr>
<tr>
<td>Claimant</td>
<td>The type of account, platform, or source responsible for posting the misinformation.</td>
</tr>
<tr>
<td>Claim Date</td>
<td>The date on which the original material was created or posted.</td>
</tr>
<tr>
<td>Review Date</td>
<td>The date on which the fact-check article was published.</td>
</tr>
<tr>
<td>Verdict</td>
<td>A one or two-word summary of the verdict reached by the fact-checking organisation, e.g. ‘False’ or ‘Partially false.’</td>
</tr>
</tbody>
</table>

Processed results were stored securely in a CSV file, and articles which dealt with non-English or irrelevant source material were manually screened out, as were those reporting on rumours (claims which have yet to be verified). This reduced the size of the sample from 301 to 191 unique fact-check articles. Each article was manually inspected, and a more detailed version of the verdict was extracted. This was usually a two or three sentence summary describing the level of falsity of the post and the organisation’s justification for giving it a particular rating. This longer verdict provided more context than the one-word rating initially returned by the API and was necessary to properly standardise verdicts across different fact-checkers which each maintained their own rating system.
4.1.2. Data Preprocessing
Each article usually included a link or screenshot of the original post containing the misinformative claim, allowing more content-specific information to be extracted. Posts were classified as belonging to one of the following three media formats: text, video, or image. Within each of these high-level categories, inductive codes were further applied to capture notable details, such as whether the post was a screenshot from elsewhere on the Internet. Posts were also coded according to whether they had been assigned a fact-checking label by the platform, had been deleted, or were still circulating with no label.

More fine-grained, but still anonymised, information about the claimant’s social media account was also processed where available. After an initial review of a few posts, account and page types were classified deductively as either belonging to: a member of the public; a mainstream celebrity or public figure; an online influencer; or a business organisation. Account types were judged by inspecting user profiles for verified badges, reading account descriptions, and inspecting a sample of recent posts where publicly available.

Finally, a typology of the misinformation in the sample was constructed inductively, due to a lack of scholarly work on women’s health misinformation, and the fact that many empirically derived typologies are specific to certain topics or health issues, such as COVID-19 [69]. On the first pass, open codes were applied to the verdicts based on the post’s level of veracity, and the ways in which the false information had been presented. For instance, completely false posts with no grounding in reality were coded as fabricated, whereas partially false posts were coded as misleading. A similar approach was used to categorise the claims made by each post: claim descriptions were coded at a low level of abstraction according to their topical focus, and iteratively grouped into higher level categories.

4.1.3. Data Analysis
After defining each of the codes and categories more clearly, the data was re-coded a final time to ensure consistency and accuracy. Then, a frequency analysis was conducted on each of the variables described above using Microsoft Excel: the occurrence of each code in the dataset was counted and noted in a separate worksheet. The full codebook for the data is provided in Appendix A.

4.1.4. Ethical Considerations
Ethical concerns associated with using social media data include the reidentification of participants through quotes and published handles, and a lack of informed consent from participants [4]. Though publicly available, users may not reasonably expect their data to be repurposed for scientific research, making it problematic to justify the use of this data as ethical simply because it is accessible [77]. As such, only metadata was collected for the social media posts used in this study, and no sensitive data was stored or accessed. High-level codes were applied at the time of accessing the post, making it difficult if impossible to reidentify posts from the dataset. The study was approved by the UCLIC ethics committee under the departmental ethics code UCLIC_1920_007_Staff_Singh.

4.2 Results
4.2.1. General Overview
A total of 191 posts were included in the sample, dated between 5th July 2021 and 27th July 2022. The average time period between a post being created and reviewed by a fact checker was 17.2 days, though there was variation in this value (SD=37.1 days). Further illustrating this variation is the finding that the longest time period between claim creation and review was almost a year (310 days), and the shortest was 1 day.

As illustrated in Figure 2, the number of posts per month remained steady across time, with the exception of a sharp peak in May 2022. Further decomposition of the data shows that 57% of posts created during the months of May and June 2022 were related to contraceptive and reproductive health—in particular, abortion access. Though causality cannot be inferred from this data, this increase in reproductive health misinformation coincides with the high-profile US Supreme Court’s decision to overturn Roe vs. Wade, legislation granting the constitutional right to abortion [74]. As will be discussed, many posts in the sample from this time period describe DIY and unofficial abortion methods, suggesting a relationship between current affairs and the focus of women’s health misinformation on social media.

![Number of posts created each month between July 2021 and July 2022](image)

4.2.2. Topical Focus and Claims
As shown in Figure 3, the largest category of posts focused on reproductive health (40%). Posts in this category included misinformation about abortion procedures, methods of contraception, and factually inaccurate claims related to pregnancy, breastfeeding, and childcare. Several US-based posts described potentially dangerous herbal methods for obtaining abortions and were often caveated with the disclaimer “avoid these herbs if pregnant.” However, the timing would suggest that they were intended as an alternative method for abortion in the wake of Roe vs. Wade being overturned [62].
The second significant category of misinformation comprising 25% of the sample was related to COVID-19 vaccinations and in particular, their alleged risks to pregnant or breastfeeding women or their negative impacts on female fertility. Cancer-related misinformation, which made up 16% of the sample, presented ineffective cures to cancer, and framed everyday household products or foods as possible remedies. Lastly, the smallest share of the sample (14%) was comprised of bogus weight loss products or regimes. Supplements, herbs, and dieting methods were presented as easy methods for rapid weight loss, and usually had little to no clinical backing.

**Figure 3: Distribution of topics across the sample**

### 4.2.3. Media Formats

An equal distribution of text, image, and video formats was observed across the sample, as illustrated in Figure 4. A smaller proportion of posts were digital news articles (13%). Videos were dominated by low-fidelity and selfie-style skits, with this style of content making up almost three quarters of video-based posts. Only a quarter of videos featured a public figure, or appeared to be professionally produced, e.g., a TV broadcast. Just over two-thirds of text-based posts were expository and provided information without an emotive slant, though a notable minority of post were politically charged or sarcastic in their delivery. The most common image format was screenshots, either of social media posts from other platforms or of technical reports and official documents. Almost all screenshots were captioned by the user, in such a way that either recontextualised or reframed the information, or ‘charged up’ the post with political or emotive commentary.

**Figure 4: Distribution of video, text, image, and article formats across the sample.**

### 4.2.4. Types

Only 20% of posts in the sample contained completely fabricated information. The remaining posts contained varying types of reconfigured, manipulated, and decontextualised facts. The most significant proportion of posts were marked as misleading (34%): that is, information which is partially true, but incomplete, e.g., suggesting that a plant which has been shown to reduce cancer symptoms, can cure the condition altogether. Similar to this category was missing context, which 24% of posts fell under. These posts included information which may have been true under certain conditions, but was presented in a different context or was omitting contextual information altogether. This included satirical posts being presented as serious, and information that was missing crucial research context, e.g., applying the results of studies done on animals or tissues to humans. The final category, comprising 22% of the sample was misrepresentation of data—a category reserved for instances of flawed scientific and numerical reasoning such as mistaking a correlation for a causation and cherry-picking statistical figures.

**Figure 5: Proportions of misinformation types in the sample**

### 4.2.5. Sources

As outlined in Figure 6, the majority of posts in the sample were on Facebook (56%), with smaller amounts coming from Twitter (13%), Instagram (10%), and TikTok (4%). A small proportion of posts originated on miscellaneous platforms including Reddit, Quora, and YouTube (6%). Only 11% of the sample originated from news platforms, both mainstream and alternative.

In addition, the analysis revealed that the largest proportion of women’s health misinformation was bottom-up, and shared by ordinary members of the public (46%). Still, as evident from Figure 7, a significant proportion originated from accounts belonging to public figures (29%). Online celebrities and influencers espousing various forms of alternative medical treatments were common, however, mainstream celebrities, journalists, and political figures were also represented in this category. For instance, a quarter of accounts labelled as belonging to public figures were linked to US and UK politicians. Lastly, 17% of the misinformation in the sample originated from an account belonging to an organisation, over half of which were businesses promoting a product. This illustrates the diverse mix of misinformation sources in the sample, where mainstream, alternative, and commercial sources saw clear representation.
4.2.6. Platform Interventions

Lastly, the results show that a sizeable portion, but not all, of posts have been subject to direct platform intervention (48%). Such interventions include the affixing of fact-checking captions or filter screens to misinformative posts and providing links to fact-checking articles. Similarly, many posts have since been removed or are unavailable (16%)—though it is unclear whether this was the result of platform moderation, or the user deleting the post. Still, just under a third of the posts were still in circulation with no fact-check warning as of 31st July 2022, as illustrated in Figure 8.

5. STUDY 2: DIARY STUDY

5.1 Methods

5.1.1. Participants

Participants were recruited on a rolling basis between June and July 2022 through purposive and convenience sampling. A pilot sample consisting of four female participants between the ages of 24-30 was initially drawn from the researcher’s social network to better define the study’s target audience and refine the method. The pilot study suggested that participants who used at least one popular social media platform daily were more likely to serendipitously encounter misinformative or provocative health content on a regular basis. As such, recruitment was focused on individuals likely to fit this criterion: participants were recruited via social media, and flyers were posted to public and private groups on Facebook, as well as on the researcher’s Instagram, LinkedIn, and Twitter accounts. Participation was limited to individuals living in the UK to ensure a baseline level of consistency in the social media and news culture participants were exposed to. The flyer used for recruitment is provided in Appendix B.

A total of 19 participants were recruited, of which two only partially completed the study. Participants were all living in the UK, spoke English fluently, and identified as women at the time of recruitment. The majority of participants were aged between 18 and 34 years (n=15), with the rest aged between 35 and 44 years (n=3) and 44 to 55 years (n=1). Participants’ highest level of education varied from A-Levels (n=1) to undergraduate (n=7) and postgraduate (n=11) degrees. All participants reported daily use of social media.

5.1.2. Materials

Screening Survey

A short screening survey was administered to participants at the start of the study via an anonymous Qualtrics link, to collect basic demographical data and information on their social media usage habits. The full questionnaire is provided in Appendix C.

Briefing Sheet and Interview

An initial briefing interview lasting approximately 20 minutes was held with participants remotely, to introduce the research and collect basic contextual information about their experiences with health misinformation. The first section involved a semi-structured discussion about participants’ habits when seeking health information online, their strategies for ensuring the accuracy and reliability of information, and their general perceptions towards misinformation on social media. Following this interview segment, participants were introduced to the diary study: a dedicated briefing sheet was shown to participants via screen share and summarised live by the researcher. The sheet was written in an accessible FAQs style and clarified details such as when and how often participants were expected to submit an entry, and the procedure for recording entries. Only the first section of the briefing interview was recorded, and the audio file was deleted following transcription. Anonymised transcripts were stored locally within an NVivo project. The topic guide for the interview and the diary study FAQs sheet are provided in Appendices D and E respectively.
Diaries

Participants completed digital diaries which were accessible via a Qualtrics link provided on the diary study FAQs sheet. A structured questionnaire with a mixture of closed and open questions was selected as the diary medium to standardise how participants reported events, and to ensure they included details relevant to the situated sensemaking process. To elicit basic contextual information, the questionnaire included multiple-choice questions asking about the source of the misinformation, and participants’ activities at the time of the encounter. Then, participants were instructed to describe the post in their own words, and why they found it provocative or misinformation. Finally, the structure of gaps, bridges, and outcomes was used to elicit a rich, step-by-step description of participants’ sensemaking process, and how it evolved as they interacted with the content. Following the pilot period, an additional prompt was added to the diary, which asked participants to detail any technical or architectural features of any of the websites or platforms accessed that they felt helped or hindered them when making sense of the content. It was hoped that this would more explicitly probe the impacts of platform design and layout on the sensemaking process. The full diary questionnaire is provided in Appendix F.

Follow-up Interview

A semi-structured follow-up interview was conducted with participants to allow reflection on the diary process, and elicit additional post-hoc information on their entries for triangulation. All interviews were conducted remotely and lasted on average, between 30 and 45 minutes. Following a general discussion about the diary process, elements of Dervin’s micro-moment timeline interview were employed: the gaps described by participants were probed using a ‘5Ws’ paradigm (who, what, when, where, and why) [21], and participants were asked to reflect on their bridging strategies and whether their gaps were resolved as expected. As before, the audio recording was deleted soon after transcription, and transcripts were pseudo-anonymised and stored locally. The generic topic guide for the follow-up interview is provided in Appendix G.

5.1.3. Procedure

Interested participants were instructed to contact the researcher via email, and were sent a link to an information sheet and consent form published on RedCAP which is provided in Appendix H. After reading and signing, they were invited to complete the pre-study survey which further assessed their eligibility. Once participants were confirmed as interested and eligible, they completed a short briefing interview before beginning the diary elicitation period.

In line with previous work [64] a critical moment diary approach was used, as opposed to eliciting moments from participants at fixed times and frequencies. This better captured the serendipitous and naturalistic contexts in which misinformation was encountered and reduced the likelihood of participants feeling pressured to artificially elicit events by actively searching for posts to record. Initially, participants were asked to record an entry whenever they encountered health misinformation on social media or any websites they visited. While the diary study FAQs sheet contained a list of relevant health topics adapted from Study 1, it was emphasised that participants were free to record posts about any health conditions or topics which were important to them.

The target number of entries was at least five over a period of two weeks, but during the pilot study, it became evident that some participants struggled to obtain enough entries and had difficulty deciding whether the posts they encountered were true instances of misinformation. Consequently, the prompts were revised to encourage a wider range of critical moments: participants were instructed to record entries when they came across provocative, exaggerated, or otherwise misinformation health content. It was hoped that this would capture not only cases where the content was obviously misinformation, but also more nuanced encounters which still caused uncertainty or intrigue.

After participants had submitted their entries, they were invited to complete a debriefing interview, after which they were sent a £15 digital One4All voucher. The two participants who did not complete the study in full were each reimbursed with a £10 voucher.

5.1.4. Data Analysis

The first stage of data analysis overlapped with the collection process, and involved transcribing, critically reading, and synthesising interview transcripts and diary entries as they came in. This immersion process revealed links and patterns between participants’ pre-study interviews, and the views expressed in their diaries—intuitions which were validated and queried further in the follow-up interviews. Once participants had completed the study, data from their interviews and diary studies was integrated, and each encounter with misinformation was contextualised and resequenced.

Thematic analysis [13] was used to analyse the processed interview transcripts and diary entries on account of its methodological flexibility and compatibility with other theoretical frameworks. First, data was inductively coded line by line at a low level of abstraction to condense the data into its key sensemaking elements. Dervin’s metatheory was then used as an additional deductive tool: identified codes were mapped to the key sensemaking dimensions of background, situational context, gaps, bridges, and outcomes, and organised into coherent themes. Where participants gave consent, they were occasionally contacted during the analysis process to clarify certain points and validate emergent links and themes as a form of member checking [9].

5.1.5. Ethical Considerations

All participants read and signed an informed consent sheet which clearly outlined the study’s procedure and the methods of data processing prior to participating. All interview segments were recorded with the explicit consent of participants, and additional verbal consent for the use of direct quotes was attained if the participant disclosed sensitive information related to their health during an interview. All data from the interviews and diaries was pseudonymised using a participant ID, and all identifiable details were removed from the transcript and quotes unless necessary to provide context. To avoid the ethical issues related to directly accessing social media posts outlined in the previous study, participants were not asked to provide links to the posts they encountered. Rather, they were asked to describe in their own words, the format and content of the material in their diary entry. As before, the study was approved under the departmental ethics code UCLIC_1920_007_Staff_Singh.
5.2 Results

First, a descriptive overview of the diary data will be given, and then, the qualitative findings will be described. The qualitative results are organised under six overarching categories: user disillusionment towards social media platforms and their management of misinformation; socio-technical contexts of misinformation encounters; sensemaking gaps; individual bridging strategies; collective bridging strategies; and sensemaking outcomes. Throughout this section, the citing convention of (PX, EZ) where X denotes the participant ID and Z the diary entry ID is used to refer to data extracts. Where the data is from an interview, the convention (PX, I) is used.

A total of 75 entries were submitted between June 1st and July 24th. Posts were overwhelmingly skewed towards a focus on dieting and body image (49%). Other topics such as mental health and neurodiversity (15%), physiological health (12%), reproductive health (11%), and public health (8%) received less, but still notable, representation in the sample. Participants decided to share only 9% of posts, and no posts were reported to the platform.

Misinformation was encountered on a range of platforms, the most common being Instagram (36%), Facebook (24%) and TikTok (16%). Twitter was used in 9% of encounters, and 15% of the encounters occurred on other platforms, including YouTube, Reddit, Discord and Pinterest. The largest share of posts originated from the accounts of members of the public who the participants did not know personally (39%), and the second largest originated from business accounts (29%). A smaller proportion of posts originated from official media and news organisations (15%), and 12% of posts originated from a public figure or social media influencer. Only 5% of posts in the sample originated from a user the participant knew personally, such as a friend or family member.

5.2.1. Disillusionment and Distrust

The first major theme in the data was a general distrust of online health information, and disillusionment with the ability and willingness of social media giants to regulate misinformation. Almost all participants expressed an inherent skepticism of health information from online sources, with exceptions being the websites of official bodies such as the World Health Organisation (WHO). Underpinning this skepticism were concerns about the low barriers of entry to content creation on social media, and the ease with which non-experts could produce medically inaccurate content.

“There’s a lot of misinformation online. Whenever I scroll Instagram or Facebook, I see a lot of people without degrees or qualifications posting advice about diets or health.” (P4, I)

Moreover, participants generally held a negative perception towards social media platforms’ role in managing misinformation, believing them to be uninterested and ineffective mediators. As well as recognising difficulties in scalability when regulating such a vast quantity of information, participants believed that social media companies were more concerned about engagement and profit than the quality of information circulating on their platforms.

“[Social media platforms] are terrible at managing [misinformation] [...] it’s all about the money for them.” (P11, I).

Some participants were aware that platforms employed steps to combat misinformation, but either felt that such features and systems were only operational on the back-end and obscured from them as users, or believed interventions to be “falling short” of the effort required to fully tackle misinformation. Some participants perceived platform intervention to be selective, and solely reserved for high-profile cases and topics. While they had seen fact-checking labels applied to misinformation surrounding topics such as COVID-19 or the 2016 US presidential election on Twitter, they had yet to see the same level of attention given to other issues or health topics. This only added to their concerns that health misinformation was spreading “unchecked and unregulated” on social media.

“Recently I met a designer who works for Meta, and he said they were using some third-party solution for fact-checking. So they have a dedicated system [...] but it’s not visible on the UI.” (P4, I)

“If you read something about coronavirus there were disclaimers saying, ‘be careful if it’s true or not’ but if you search for other things like menopause or cancer or diabetes, there’s no disclaimers.” (P14, I)

Lastly, many participants expressed concerns that platforms could not be trusted to responsibly combat misinformation while respecting the free speech of users. Some participants were wary of moderation algorithms they perceived to be imprecise and punitive, and expressed frustration at false positives, where content which was not misinformation had been flagged as such. These participants believed strict content regulation to be tantamount to censorship, and felt added frustration at their inability to correctly predict the outcomes of black-box moderation algorithms. Here, there is some evidence that users of Facebook have adjusted to the perceived threat of being banned from the platform by either setting up alternative profiles, or avoiding keywords known to trigger moderation algorithms.

“People always get blocked by the algorithm who don’t deserve it [...] I’ve got a friend who has something like three alt profiles set up just in case she gets kicked off one.” (P11, I)

“With the algorithm, you could only say one word and then you’re banned [...] people have had to type certain words in a different way or come up with a code word to avoid it. It’s unfair as we’re all entitled to our opinions.” (P12, I)

5.2.2. Socio-technical Contexts of Encounters

The architecture of social media platforms played a significant role in how and why misinformation was presented to users. Three key structures were found to configure participants’ encounters with misinformation: networks of followings and friendships; closed communities; and algorithmic content recommendation.

Social Networks

As mentioned previously, only a minority of misinformation was identified as originating from an individual the participant knew in real life. However, participants’ social networks still played a role
in shaping their reactions to misinformation. Participants generally maintained a strong mental model of the people they followed on social media, and therefore, had a sense of the content they should expect to see. Content which did not fit the topical and stylistic pattern of postings for a particular account or individual was often flagged as suspicious, and subject to more critical thought and skepticism than posts which were ‘run-of-the-mill’.

“You kind of know who you follow and what you should get in your feed. So if it doesn’t make sense in terms of the information that you normally get, you know that it’s an advert” (P19, I)

“I found this post thought provoking because I wasn’t expecting it from [celebrity] [...] He usually posts photos of what he’s doing in his daily life, or he promotes his music.” (P9, E4)

Source-based credibility heuristics were the most compelling in contexts where the sources of information were personally meaningful to participants, such as when it was a person they knew, or a verified account they followed. Many participants took content at face value if it originated from people they perceived to be trustworthy, such as verified journalists and broadcasters. Similarly, knowledge of what was expected from a particular individual or information source was observed to prime participants for misinformation. Participants often automatically dismissed information from sources they perceived as unreliable, such as poor-quality news outlets and individuals who had a habit of posting misinformation.

“I initially found [post] trustworthy because it was from a podcast host I follow” (P2, E5).

Gated Communities

Several participants encountered health misinformation within closed or curated communities, such as private groups on Facebook, discord servers, and subreddits with strong moderation. The closed communities described by participants were often specialised in a specific health condition. As such, participants often interpreted the misinformation they encountered in light of these conventions, reacting in ways they perhaps would not have had the information been encountered elsewhere. For instance, one participant who came across an image on the subreddit ‘r/instagramreality’, a Reddit community focused on documenting and critiquing heavily edited Instagram selfies, believed that without the context of the subreddit, she would not have given the image much attention and would have assumed it was authentic. In this case, the migration of the image from one social media context to another led to a different emotional reaction to the material.

“When it’s on r/instagramreality, and you know it’s there because something in the image is fake, it’s a whole different experience [...] it made me mad that [the poster] felt she had to alter her own image even though she is really fit and muscular.” (P17, E1)

Moreover, many groups primarily catered to women, producing added skepticism and reservation towards male participation that was perceived as either not belonging in the community, or domineering. For instance, one participant became “annoyed” at a user she identified as male promoting a bogus cure for polycystic ovary syndrome on a support group for women living with the condition (P18, E2). To an extent, this attitude extended beyond closed groups, and into more traditional, unmoderated social media contexts such as Instagram comment sections on posts about menstruation which participants believed ought to prioritise female participation. As such, the perceived purpose of a digital space, regardless of whether it was formally moderated, affected how participants framed the information originating from there.

“I got increasingly annoyed because then men started leaving comments about how the pain of getting kicked in the balls is worse than period cramps, and this clearly isn’t a post for them.” (P17, E2)

Algorithmic Content Recommendation

On many occasions, participants encountered misinformation on their feeds or ‘For You’ pages that was not from a group nor from anyone in their friends or follow list. As such, participants frequently reflected on the reasons why they were being targeted with misinformation and constructed folk-theories to explain the operation of the content recommendation algorithm. Often, they used language that was highly relational and even animistic at times, imbuing the algorithm with a sense of intent and purpose. Some participants were able to point to a specific earlier browsing session as the catalyst for the algorithm recommending certain content even when the event occurred on a different platform, assuming some obscure, Internet-facilitated connection between them.

“I came across a video about this a few days ago [...] I think because I clicked that video about it, the Internet advised me similar content on a different platform, so there must be some connection.” (P8, I)

Similarly, even when participants could think of no obvious prior event which led to the content appearing on their feed, they were able to connect their general interests and pattern of account followings to the topic of the misinformation being shown to them. One participant who came across reproductive health misinformation on Instagram, while disapproving of the content, felt that she could “sympathise” with the algorithm for recommending it, on account of its consistency with her usual pattern of engagement with posts focusing on women’s issues (P17, I). Some participants also believed that certain posts were targeting them on account of their gender. This was usually the case with weight loss misinformation, where the visual imagery of posts or advertisements was obviously gendered.

“I think it specifically targeted women, who naturally have rounder tummies from their wombs and the pictures of the people demonstrating the exercise were women.” (P6, E2)

However, some participants felt their autonomy and privacy was degraded by the perceived omniscience of algorithmic recommendation. This prompted some participants to tailor their account
settings such that content on their feeds only originated from accounts they followed or from sources they trusted. Motivations for doing so were mainly centred around a desire to regain autonomy over their information consumption, filter out unwanted or irrelevant content, and to mitigate the perceived risk of radicalisation by recommended conspiratorial content.

“You feel like you’re being watched by the Internet. Is this really the content I’d search for if I was using the platform by myself, or is it just what the platform wants me to see?” (P8, I)

“I’ve heard a lot of people that started believing in weird theories because the YouTube algorithm starts pushing things onto people, and that’s quite scary [...] so I have deactivated autoplay.” (P2, I)

5.2.3. Sensemaking Gaps
The gaps associated with misinformation encounters broadly fell under three categories: content-specific queries, moralistic and safeguarding concerns, and uncertainties about personal health.

Content Related Gaps
Content-centred gaps were either related to the accuracy and medical safety of claims being made in a particular post, or a post’s intended message, i.e., what it was “trying to say”. Many posts either lacked detail or made exaggerated claims which were at odds with participants’ existing knowledge. Weight loss misinformation was especially perplexing, and many participants found themselves questioning the safety of methods promising unusually fast weight loss through unconventional techniques. Concerns about medical efficacy and safety extended to other issues besides weight loss: topics ranging from blood pressure monitoring to sleep health were flagged as suspicious on the basis that they were lacking in key details and endorsements from medical bodies.

“Losing 24 pounds in less than a month sounds horribly unhealthy [...] [metabolic diet] doesn’t sound very science-based.” (P7, E2)

In some cases, participants were unable to understand the message of the posts they came across, a lack of comprehension which left them confused and frustrated. These posts often used vague language and confusing imagery to grab participants’ attention, a tactic that one participant referred to as “provocative” and “annoying”, in the context of an Instagram reel which gave ambiguous advice about how to lose weight (P20, E2). Here, gaps were less related to the veracity or safety of a particular claim, but focused on the underlying intention or narrative being conveyed by a post.

“I don’t understand what this post is trying to convey about period pains. It made me question why a post like this is so popular if the message it’s communicating isn’t clear at all.” (P17, E2)

Moralistic and Safeguarding Concerns
One striking dimension of the encounters was the moral disapproval they evoked, and in particular, the strong sense of safeguarding that participants expressed towards groups they perceived as vulnerable, such as younger social media users. Examples of moralistic concerns included strongly worded disapproval of medical misinformation which had the potential to harm others, particularly when it was created with financial aims. News outlets which used fearmongering headlines to describe public health issues were also perceived as irresponsible, and were believed by participants to be contributing to anxiety in the general public. Participants perceived people to be more emotionally vulnerable in the wake of the COVID-19 pandemic, as they had “lost family and friends”, and disapproved of news outlets leveraging that fear for “clicks” (P2, E2).

“The intent is to do with money and they don’t care about how they’re going to hurt somebody or even kill somebody and that makes me really angry.” (P11, I)

Lastly, some participants expressed dismay at the possibility of younger, more impressionable girls coming across, and being negatively impacted by weight loss misinformation. They believed this content to be harmful and often related it to their own experiences with body insecurity in adolescence. These memories were often experienced viscerally, and added to their sense of concern about younger users being misinformed by reductionist information about the ideal body image for women.

“Young girls whose bodies haven’t even started [developing] will see this [...] from the age of 10 to 20 I was annoyed about my body shape [...] so these posts just make me a bit sad” (P17, E3)

Questioning Health
Misinformation caused participants significant consternation when it was at odds with self-knowledge they were previously confident in. The example of neurodiversity-related misinformation on TikTok is used to illustrate this point. Many participants reported coming across content on TikTok that oversimplified conditions such as ADHD and dyslexia down to a small number of behavioural traits, presented in a bullet-point style. Often, these traits were vague and arbitrary, and not specific to the conditions presented. Despite not having a formal diagnosis and never previously entertaining the possibility, some participants wondered whether they were in fact neurodiverse, causing them self-doubt and confusion. Though participants in this category eventually decided that they were unlikely to be neurodiverse, they felt that a significant amount of time had been spent wondering otherwise and expressed concern about the possibility of “teenagers self-diagnosing” (P7, I) on the basis of these posts—indicating the formation of a safeguarding concern.

“They posted some ridiculous video saying something like “You are neurodivergent if you do this this, and this”, and I was like hold on a sec, I tick a few of these boxes [...] I got so sucked into it.” (P10, I)

This style of content also caused anxiety for participants with neurodiverse conditions. One participant with a formal diagnosis of dyslexia described coming across a video which presented a “reductionist and annoying” view of the symptoms of dyslexia (P5, E3). This caused her to reflect on her diagnosis process, believing videos like this to have been unhelpful and anxiety-inducing, rather
than enlightening. The same participant believed that the brief style of TikTok videos was a factor underlying the reductive nature of neurodiversity-related content on the platform. Here, the convention of informal, short-form content dissemination was thought to encourage brevity over nuance, resulting in oversimplified portrayals of complex conditions.

“I knew it was going to be a quick 30s video reducing things down [...] it reduces a complex learning difference down to five signs” (P5, E3).

5.2.4. Individual Sensemaking Bridges

Often, participants were able to effectively overcome their uncertainties individually, by drawing from internal knowledge or critically synthesising additional sources of information where necessary. Three key individual bridging strategies are described here: inferring poster intent, critical reading, and secondary online searching.

Inferring Intent

As mentioned in section 5.2.2, the source of posts influenced participants’ views of the information contained within. Where participants were unfamiliar with the source, they often introspected on the intent of the poster as a means to judge information credibility. If they inferred a financial or self-promotional intent through the nature of the account or the framing of the information, they became far more skeptical of the content and were more likely to dismiss it as exaggerated or misleading. Underlying this was the assumption that businesses or individuals promoting a product would be more likely to lie or exaggerate information as a persuasive technique. A similar line of thinking emerged in relation to “clickbait” headlines. While not directly promoting a product, participants believed news outlets to be financially motivated, and applied similar assumptions to headlines which appeared to be more concerned with attracting their attention than providing balanced, accurate information.

“It’s an advert from a business wanting you to purchase their products so I wouldn’t ever buy something like that without digging deeper and doing my research.” (P11, E4)

“[Headline] is just for clickbait and views, the media is running a business.” (P12, I)

There is some evidence that UI elements aided participants in making rapid inferences of creator intent, and by extension, judgements of credibility. Banners and visible indicators that a post was a sponsored advertisement immediately signposted to participants that a post had been created with a financial or promotional motive. This removed the burden of inferring intentionality from the participant and usually made them less trusting of the content and more comfortable dismissing it. For instance, one participant cited a tendency to “automatically skip” Instagram posts which were labelled as paid promotions (P1, I).

“There was text that said ‘sponsored’ under the post [...] It makes me a bit more skeptical and less likely to trust it as it just comes from a business.” (P14, E3)

Critical Reading

Many participants chose to read, re-read, and apply a critical lens to suspicious information. The most common activity described by participants was scanning for scientific references, citations, or endorsements from medical professionals. Participants maintained a high level of trust in official health institutions and tended to view content as more reliable if it appeared to have the backing of a medical professional. If they found no medical or scientific evidence, this was often enough for participants to judge the information as untrustworthy and dismiss it.

“I spent a minute or so scanning the article looking for anything that presented as remotely reliable or backed up by [scientific studies] but the whole article came across as clickbait.” (P12, E2)

Many participants took citations, links to journals, and screenshots of scientific figures as a sign that the poster knew what they were talking about and had done research into the topic. In one case, this led a participant to initially believe a Tweet about COVID-19, though she later discovered it was misinformation after reading comments debunking the poster’s claims. As was the case with source-based credibility heuristics, this example indicates the potential dangers that users face from misinformation which appears reputable on a surface level.

“They gave a real-life example, they used scientific terminology AND included a journal article related to facial nerve palsy and COVID-19, so I was like oh okay that seems believable.” (P20, E1)

In other cases, reading past alarmist previews or headlines was enough to assuage participants’ worries. One participant who had come across an alarming Twitter thread summarising an article about a Polio case in the UK was immediately reassured when she read the full thread, and found that the headline was exaggerated (P16, E2). In this case, the brevity of Tweets was helpful, as it allowed the participant to digest the salient points of the article quickly. She cited the convention of numbering Tweets in a thread (e.g. ‘1/5’) as useful, as it helped her “balance effort with information utility” and judge the expected length of time she would have to spend reading. A similar sentiment was echoed in relation to a provocative TikTok video about ageing: the expectation of short, and easy to digest content prompted the participant to watch the full video and get the context she needed to make a balanced decision about the accuracy of the claim.

“It was about 30 seconds long, so it was really quick to watch, and it was just a summary of a scientific paper [...] the limitations are yet to be examined for this study.” (P15, E3)

Secondary Searching

When participants were unable to judge content veracity on their own, Googling was a quick way by which they achieved clarity on content-related questions. Several participants were able to fill their information gaps with a single Google search: such was the case when the top result was from a trustworthy source such as the NHS website, or the website of an official product or brand. Participants
were able to easily digest information from these sources, as they were often written in very plain language, and used visual hierarchy to draw participants towards important information. This often led to a stable outcome, where participants were able to reach a confident conclusion about a content-related gap.

“At the top of the page there was a clear concluding statement that there was a very low risk of getting monkeypox in the UK for now, so that was reassuring.” (P1, E2)

However, if none of the first-page results were websites they recognised as trustworthy or familiar, participants were not always immediately sure of which result to click. Here, some participants employed a technique of lateral reading. This involved opening multiple results in succession and scanning them rapidly to judge which website was the most reliable or useful, often basing this decision on the presence of references to medical journals.

“There are some blog posts that look like scientific articles but are a bit old (2012-2013) […] I ended up looking at [Healthline] which had links to sources after every claim […] that provides reassurance about the quality of the information.” (P7, E1)

5.2.5. Collective Sensemaking Bridges

Sometimes, participants used the views of other users in the comments’ section to make sense of misinformation inter-subjectively. These strategies had both positive and negative outcomes, which are summarised here.

Using Comments to Make Sense: Advantages

Many participants expressed in their diaries and interviews, that they looked at comments by default when using social media. Reasons for doing so included entertainment value and curiosity about what others were saying, particularly when material was controversial. Other motivations were based around drawing from the ‘knowledge of the crowd’: sometimes participants were able to intuitt that content was misinformation, albeit for reasons they could not properly explain. As such, they instinctively scrolled down to the comments section, and skimmed what others were saying in hopes of finding an articulation of why it was misinformative.

“I’m always interested in seeing people’s reactions and comments in response to threads because sometimes it can be funny.” (P16, I)

“I guessed that my opinion would have a lot of support, so I read the comments just to check whether my expectations were right or not.” (P8, E2)

Comments were cited as being easy to digest, and often contained additional information and context about the post which confirmed participants’ intuitions and helped them understand content more holistically. Useful information included scientific and medical evidence backing up a particular perspective, anecdotal accounts and personal experiences with a particular health issue, and contextual information about the creators of misinformation, which was found to foster sympathy towards vulnerable individuals.

“The replies to the tweet were saying the woman had lost a child before, and it turned her into a bit of a conspiracy theorist […] to make fun of someone like that just felt too dark.” (P2, E3)

Comments were of particular use in helping participants overcome moralistic and safeguarding concerns, as on several occasions, the same concerns were mirrored by other users. Seeing that others were expressing similar thoughts and feelings made several participants feel validated and more confident in their opinions. This was the case for one participant who encountered body-image related misinformation on a subreddit mainly frequented by women. In this case, a sense of camaraderie was fostered with the other people on the subreddit who “knew what normal bodies should look like.” (P17, E1) Even in the absence of absolute factual certainty, participants were comfortable being led by others, and following the general ‘bandwagon’ of what other users had concluded, provided it matched their initial intuitions.

“I read through some comments, people weren’t convinced, people were confused […] everyone was saying they would [ignore the post], so I thought yeah, same.” (P20, E4)

Using Comments to Make Sense: Disadvantages

Sometimes, synthesising comments turned out to be unhelpful. When participants could not infer a consensus from the views expressed in the comments or felt that the comments did not validate their positions, they were confused and frustrated. Sometimes, participants lacked the subject knowledge to decide which commenter was correct and resorted to strategies such as scanning for scientific credentials and judging the neutrality of comments to decide who was the most trustworthy. However, this rarely bore fruit due to the informal and argumentative nature of many comments, and the fact that many users had private profiles, making it impossible to check their bios for subject expertise (P7, E2).

“I just wanted to browse, to see what people were saying, but there were just different random comments about different things so it didn’t sway me one way or another. I didn’t come to a conclusion.” (P2, E2)

Furthermore, some participants described getting “sucked into” the comments, to the point where they became distracted from their original questions about content accuracy. In such cases, the entertainment value of the comments became a barrier to attaining clarity, and simply pulled them along rabbit holes of reading comment threads and spectacleing “comment wars.” Sometimes, participants consulted comments expecting other users to agree with them, only to find the opinions being espoused offensive or lacking in empathy. This often led to the spawning of new gaps, particularly moralistic ones, and ended with participants disengaging from social media with their original questions unanswered.

“There was no concern about people who can’t breastfeed […] but formula is still a good way [to feed infants]. It was demotivating to see those comments, because they were all just hyper fixating on “yay I can breastfeed” (P20, I)
5.2.6. Actions, Decisions and Outcomes

Having discussed the intermediate outcomes of various bridging strategies, attention is now paid to the termination and overall conclusions of participants’ encounters with misinformation. A few overarching categories were observed: reinforcement of a particular viewpoint, or disengagement from content due to disinterest, emotional burnout, or anxiety. Furthermore, participants rarely decided to report or share the content they encountered. These decisions are also discussed in this section.

Moving On

Many participants chose to stop engaging with posts once their interest waned below a certain level. This was usually connected to a disinterest in the topic, or informational saturation: that is, they were confident in their current frame of understanding and believed that further investigation or engagement would only confirm what they already believed to be true. Even when participants were not totally certain that a post was misinformation and believed there could be some truth to the content, they dismissed it if the information was incompatible with their lifestyles, habits, or personal observations. In such cases, participants prioritised their existing frame of understanding and felt no pressing need to know if a post was “100% factual”, provided the topic was not important to them.

“I stopped looking at [the post] because I was pretty sure it was nonsense.” (P12, I)

“So maybe there is part truth to [post claiming you should not sleep with a pillow] but I just don’t want to admit it because I like sleeping with five pillows [...] even if it was true I would dismiss it” (P10, E3)

Information Avoidance

Sometimes, the emotional weight of certain misinformation made it difficult for participants to keep engaging with it, even if their questions had not been resolved. As discussed in section 5.2.5, negative emotions such as frustration, sadness, and anger often resulted from prolonged exposure to upsetting comments, leading participants to disengage from social media. Emotional burnout also resulted from exposure to posts which dealt with “heavy” or “dark” topics such as infant mortality and medical discrimination. Even when an information gap related to the content had been overcome through critical reading, emotional angst remained, which in the case of some participants, was not resolved except by the passing of time.

“I clicked and read the article but didn’t do more than that, reading the article just made me sadder and angrier.” (P2, E4)

Similar patterns of information avoidance were observed in participants who did not want to verify the accuracy of social media posts, for fear of coming across information which would upset them. These participants preferred to remain uncertain about the specific details of a post or article, as they judged the subject matter too sensitive to research further, and likely to be distressing.

“I didn’t research [news article] further because it was quite sad...I didn’t really want to know what they’d done in more detail.” (P12, I)

Sharing and Reporting Decisions

Though all social media platforms provide an opportunity to report content, no participant in the study chose to do so. Strongly influencing this decision was the fact that almost all the misinformation encountered by participants was nuanced and not wholly false, making the act of reporting the post inappropriate. Despite not explicitly being familiar with the community guidelines of the platforms they were using, many participants believed that reporting was reserved for “truly outrageous” or illegal content. The posts participants encountered were often subjective or subtly misleading, and many were advertisements which were not deemed harmful enough to warrant violation of the platform’s guidelines.

“I don’t think it’s actually scamming anyone, it’s just another business that really wants you to pay for their app.” (P3, E1)

“It’s not illegal and there is a sea of content like that, me reporting one wouldn’t make much difference.” (P6, E2)

Furthermore, while a minority of participants shared the content they encountered with others, the vast majority did not. Most of the time, the content was simply deemed uninteresting or not relevant to anyone the participant knew. Other justifications were more moralistic in nature, and centred around not wishing to upset others, particularly when the content touched upon sensitive topics. Other participants were wary of sharing information they felt could be misleading, from a perspective of not wanting to misinform their friends. One participant expressed a reluctance to share health-related content on social media in general even if it was accurate, in case it did “more harm than good” (P11, I).

“I didn’t share because I don’t know how reliable the information is and I wouldn’t want to spread misinformation and worry to my loved ones.” (P20, E1)

6. DISCUSSION

The first study in this dissertation aimed to characterise the types, sources, and claims of women’s health misinformation on social media, and critically appraise the extent to which social media platforms have intervened in its spread. The second study investigated women’s encounters with health misinformation on social media using Dervin’s situated sensemaking methodology as a framework for understanding user cognition. Overall, the findings across both studies reveal a great breadth in the sources, formats, and types of misinformation women are exposed to, with Study 2 in particular, illustrating how these differences can configure specific emotional and behavioural outcomes for participants. Participants displayed a tendency towards inter-subjective sensemaking, and described how specific social media features both helped and hindered them when processing misinformation. Lastly, the studies paint a mixed-picture of platform intervention: while some evidence for timely fact-checking of women’s health misinformation was observed in Study 1, this was limited, and garnered a negative perception from participants. These findings will now be discussed in more depth, in relation to existing literature and in terms of their potential for informing novel, community-based fact checking interventions.
6.1 Types, Sources, and Claims of Women’s Health Misinformation

6.1.1. Topical Focus
The results of Study 1 suggest that posts about reproductive health dominate the landscape of women’s health misinformation on social media. However, this trend may be explained by the priorities of individual fact-checkers, rather than being representative of the wider universe of misinformation online [69]. Many posts surrounding reproductive health were published between May and June 2022, and primarily focused on abortion and contraception. As mentioned, this period coincided with the overturning of Roe vs. Wade, and though further work is required to confirm this intuition in relation to women’s health, high-profile and topical content may be more likely to receive attention from both fact-checkers and platforms, as was the case for COVID-19 misinformation during the height of the pandemic [69]. Still, these results are of interest, as existing work has found an abundance of anti-choice abortion misinformation online [32]. On the other hand, Study 1 revealed the existence of pro-choice misinformation geared around self-managed, herbal abortion in the face of increasingly restrictive abortion access in the United States [74], suggesting diversity in the ideological framing and intent of reproductive health misinformation.

While weight loss misinformation was found to receive the least attention from fact-checkers in Study 1, it dominated the misinformation that participants encountered in Study 2. Some participants suspected that they were being targeted by this content on account of their gender, and believed that advertisements were promoting a particular body image rather than health. This is consistent with results from Dedrick et al. [20], who found a gendered slant in Pinterest-based weight loss advertisements and noted their visual emphasis on slimness and conventional beauty standards. It is known that women are disproportionately more likely to be recommended and negatively impacted by weight loss-related paid promotions on Facebook [1], and though it is impossible to determine whether participants were algorithmically targeted by gender during the study, weight loss misinformation was found to elicit the strongest emotional responses from participants. As such, the prevalence and tendency of weight loss misinformation to fly under the radar of fact-checkers is concerning, as it may pose unique risks to women’s wellbeing.

6.1.2. Sources
In line with findings from Shahi et al. [67] and Kouzy et al. [45], the largest share of misinformation was bottom-up, i.e., spread by ordinary members of the public, in both Study 1 and Study 2. Also consistent across Study 1 and Study 2 was the finding that top-down sources, while not the majority, do still generate a notable proportion of misinformation, and that these sources are occasionally high-quality or mainstream news outlets. Of interest here are findings from one US-based study, which found that exposure to deliberately fabricated news stories (i.e., disinformation) is rare, and misunderstood or misrepresented information from otherwise trustworthy news sources is a more common source of confusion [2]. As such, care must be taken not to uncritically dichotomise sources into trustworthy and untrustworthy, as traditional high-quality news outlets still present a small, but real, vector for misinformedness.

6.1.3. Media Formats
One interesting finding to emerge from Study 1 was the prevalence of screenshots in the sample, both of posts from other social media platforms, and of technical documents. This sharing modality is of particular interest, as it facilitates cross-platform migration of misinformation, and affords persistence, in that it enables posts to continue circulation even if the original is deleted. Furthermore, screenshots were virtually always captioned by users in political, emotive, or sarcastic styles. As well as having the effect of reframing and recontextualising information, additional commentary is significant, as one experiment has found that misinformation which receives user commentary when shared tends to spread faster than non-mutated misinformation [85]. Consequently, screenshots and captioned misinformation may pose unique risks both in terms of its potential to bypass content moderation, and its spreading potential.

Many posts encountered in Study 2 had a tendency towards brevity, which was sometimes helpful in allowing participants to digest content rapidly. Indeed, a robust finding is that shortened messages can be more effective, and length limitations may encourage users to express themselves more concisely [29]. The benefits of brevity have also been demonstrated for video media: shorter videos on TikTok have been found to communicate scientific concepts more effectively to a wider audience and allow individuals to receive information at accelerated rates [30]. However, brevity was occasionally found to come at the cost of information quality. This caveat is expressed by Gligoric et al. [29], who additionally found that the shorter a message becomes, the more difficult it is to maintain information quality and completeness.

The impact of brevity on information quality has been researched in-the-wild to some extent, with several papers finding that Twitter’s decision to double character limits from 140 to 280 in 2017 led to more civil, higher quality discussion on the platform [10, 39]. However, this study explored the context of political discussions, and was limited to text-based content. The current study would suggest that constraint affordances in video-based mental health messaging may lead to oversimplification and a ‘watered-down’ portrayal of complex conditions, suggesting the value of future work which explores this phenomenon in a wider population.

6.2 The Role of Social Media Platforms

6.2.1. Observed Interventions
Both studies illustrate limitations in the timeliness and coverage of fact-checks on social media posts. Even though the majority of posts in Study 1 had been either deleted or fact-checked by the platform, almost a third were still circulating unflagged at the time of the study. This is consistent with existing literature which finds that most, but not all misinformation flagged by fact-checking organisations is acted on by platforms [69]. However, no participants reported seeing a fact-check label on any of the posts they encountered in Study 2. Though this may suggest discrepancies in coverage, Study 1 shows that there is sometimes a large period of time between a post being created and being fact-checked by a platform. Therefore, it is possible that the posts encountered by participants have since been flagged.
Interestingly, a few of the claims encountered by participants appeared in the Study 1 sample with a fact-checking label, however the iterations encountered by participants had no such warning. Hence, while the first or main post identified by the fact-checker may have been acted on by the platform, different versions may not have received equal attention. This supports previous research which draws attention to the difficulties in scaling content-specific fact-checking across platforms, where posts can be quickly reproduced and reconfigured [49].

Furthermore, users had overwhelmingly negative perceptions of platform mediation in misinformation. Some believed that social media should intervene more, whereas others perceived their current intervention to be excessive, and tantamount to censorship. This is consistent with findings from a US-based study [64] which despite not stratifying by gender, lends credence to the divisiveness of platform intervention in misinformation. This paper finds an almost even split between users who think platforms should take more responsibility, and users who perceive interventions as paternalistic and punitive [64]. While the reasons for these divisions were not analysed in depth by the current study, previous surveys have illuminated an effect of political orientation on perceptions of platform responsibility [42]. This survey is again, limited by an exclusive focus on the US political climate, but illustrates the value in controlling for political orientation in future studies.

6.2.2. Algorithmic Justice and AI Governance

One surprising, yet valuable, finding was the lack of ‘mechanical neutrality’ in how participants spoke about moderation and content recommendation algorithms. They often discussed algorithms within the wider context of the commercial aims of social media platforms, and devised a wide range of folk theories to explain their underlying operation. Consistent with existing qualitative work [22], participants’ mental models varied from abstract understandings of the algorithm as a relational entity that “knew” them, to top-down operationalised theories which framed algorithms as processes with decision criteria which participants could manipulate by behaving in a particular way, e.g., avoiding language that would trigger bans or post deletions.

This raises critical questions about power in socio-technical systems and AI governance. The principles of “clarity of norms, and consistency of enforcement” [65] underpinning good governance may be violated both in terms of how misinformation is inconsistently managed by platforms, and in the lack of transparency in algorithmic moderation and recommendation. Though some elements of governance are transparent, such as public-facing community guidelines which outline policy for reporting posts, a lack of standardisation has been found across platforms both in the reporting modalities offered to users, and how reports are managed ‘behind-the-scenes’ [19]. Indeed, few participants in Study 2 maintained an accurate idea of what constituted a violation of platform guidelines, so were not confident reporting content that was not illegal or objectively harmful. Thus, the autonomy of women, as it has been in discriminatory medical contexts, may be degraded by algorithmic systems which target them with harmful misinformation, lack clear and consistent pathways for reporting content, and leave them struggling to “control their digital futures” [65].

6.3 Making Sense of Misinformation

6.3.1. The Situated Sensemaking Methodology

Dervin’s sensemaking methodology was a useful tool in framing participants’ encounters with misinformation. As well as helping to identify common sensemaking dimensions across participants, the framework made it possible to conceptualise sensemaking as a recursive journey experienced by each participant [58]. The framework’s explicit focus on situational factors made clear the socio-technical contexts of encounters, and how these shaped both the range of bridging strategies available to participants at a given moment, and their outcomes. Overall, sensemaking gaps were either informational, or moralistic and emotional. Gaps were not mutually exclusive: participants experienced several types of gaps simultaneously, or spawned new gaps as they engaged with information further.

A range of internal and external resources were drawn upon to bridge these gaps, with mixed effectiveness: sometimes they worked as expected, but sometimes, participants were left in a state of uncertainty and confusion. Lastly, several motivational and affective factors shaped participants’ decision to eventually terminate their interactions, including a lack of interest, information saturation, and emotional burnout. This spectrum of engagement with misinformation is observed in other studies [28, 78], with interactions ranging from immediate dismissal and some peripheral inspection of the source account, all the way through to lengthy research sessions involving multiple sources.

6.3.2. The Role of Affect and Empathy

While some work explores the emotional impact of health misinformation in the context of the COVID-19 pandemic [36, 43], user-centred misinformation research tends to focus on informational gaps and how these are overcome through fact-checking [23]. The limitations of such an approach are made evident by Study 2, which found that emotional gaps often superseded questions about factual accuracy. A notable finding was the high degree of concern that many participants expressed about the social impact of misinformation, particularly on those they perceived as vulnerable. Many surveys have suggested a higher degree of concern about misinformation among women [3, 61]. This study complements these findings by exploring how specific concerns manifest in real-time as a response to different types of misinformation, rather than as post-hoc rationalisations or generalised feelings. For instance, participants’ concerns about younger women being influenced by weight loss misinformation were visceral, and related to their own experiences with body image issues in adolescence.

As described in the results, only one or two participants chose to share the content they encountered. When the content was not simply uninteresting, participants expressed a fear of upsetting, misleading, or provoking friends and family if they shared it, suggesting the role of empathy in decisions to share misinformation. The extent to which empathy differs among men and women has been found to depend on the methods of measurement with the largest differences observed in self-reported data [5]. This is thought to be influenced by gender-role stereotypes, and the idea that sensitivity and compassion are feminised social expectations, making women more likely to present themselves as concerned about others [7].
It is also interesting to note that for most of the encounters, participants were using their personal social media accounts, meaning that their potential audience consisted of friends, family, and acquaintances. It has been theorised that when one’s potential audience consists of family and friends, patterns of interaction with controversial content may be subbed for fear of social backlash [75]. An interesting area for further study, therefore, may be the impact of anonymity and perceived audience on women’s decisions to share misinformation, and the interplay between social accountability and empathy in this context.

6.3.3. Social Context and Inter-subjectivity

Consistent with many studies [25, 28, 47], participants made use of source cues to judge information credibility. However, Study 2 revealed more multifaceted insights about the relationship between users’ social networks and their use of source cues. Since so few posts originated from friends of participants, it is difficult to judge whether the results of Study 2 support the existing finding that people are more likely to trust content originating from friends [25]. However, preliminary observations suggest a more complicated view of misinformation originating from familiar individuals: rather than being more likely to trust content originating from friends or people they followed, participants were mainly influenced by their expectations of a particular individual, and used mental models of their social networks to appraise the content appearing on their feeds. Thus, even when the poster was an acquaintance, friend, or family member, participants distrusted content if the individual had a habit of posting misinformation.

Furthermore, participants made strong use of comment sections when navigating health misinformation, consistent with studies which highlight people’s tendency towards inter-subjective sense-making in irregular situations [54, 73]. Existing research tends to view comments exclusively as a tool for fact-checking [66], whereas Study 2 highlights a range of motivations for consulting comments once exposed to misinformation, including entertainment value and emotional validation. Though there were downsides associated with using comment sections as a tool for fact-checking and emotional regulation, synthesising the views of other users was useful to participants when they could infer a clear consensus.

6.4 Design Directions: The Potential for Crowdsourced Fact-checking

The tendency for participants to utilise social media comments as a bridging strategy begs the question: could community-based fact-checking work in practice? Existing work suggests a mixed picture, with some experiments suggesting that crowdsourced fact-checks correspond strongly with verdicts from official fact-checkers [53]. However, previous crowd-based fact-checking initiatives such as WikiTribune encountered difficulties when deployed in practice, and were found not to scale effectively [8]. More recently, Twitter’s Birdwatch feature was piloted as a community-driven approach to identifying misinformation on the platform, allowing users to flag content they believe to be misleading and add textual notes that provide context, corrections, and sources [18]. Users can also mark notes as helpful or not helpful, and see the ratings of other notes. A recent experience sampling-based evaluation of Birdwatch found that users were more likely to rate notes as helpful if they provided links to sources, were clear, informative, and empathetic. By contrast, notes which lacked sources, were biased, and appeared argumentative were more likely to be voted as unhelpful [55].

The attitudes expressed by participants in Study 2 towards comments were similar: participants valued informative, balanced comments which gave them novel insights, and were put off by unempathetic comments or multiple users arguing. Notably, a key pain point of participants was not knowing which commenters to believe in cases where information was conflicting, as they did not tend to cite sources, and participants found that a high number of comment likes did not always indicate information quality. Birdwatch, with its credibility-focused rating system, may better facilitate judgments of information quality than the current configuration of post comment sections, as users may be able to meaningfully interpret note ratings as corresponding to trustworthiness. Moreover, given that Birdwatch notes lack conversational affordances such as replies, the likelihood of users being distracted by comment wars may be lower than in a traditional comments section.

While the feature focuses primarily on fact-checking and information gaps, it is possible that Birdwatch may ameliorate at least some of the problems encountered by participants when using the comments to judge the veracity of social media content. However, the feature has currently only seen deployment to a small set of US-based users [18], suggesting that the utility and scalability of community-based fact-checking remains an open question. In any case, this dissertation offers a user-centred glimpse into how women embark on collective sensemaking, and presents findings which can demonstrably be used to critique and inspire future interventions.

7. LIMITATIONS

7.1 Study 1

The sample of fact-checked posts was neither comprehensive nor unbiased: only the main post linked by a given fact-checking article was included in the sample, meaning that further iterations of the post were excluded. The sample was also small and limited to only a few social media platforms which receive attention from mainstream fact-checkers: misinformation from sites such as Reddit, which has a growing female userbase and Pinterest which is female-dominated [20] were rarely encountered. As mentioned in the methods section, care was taken to include as broad a range of keywords as possible when searching the API, however the search inevitably imposed topical constraints on the results returned. A traditional association between women’s health and reproductive issues may have also meant that articles tagged with ‘women’s health’ were limited to reproductive and contraceptive health.

In future, methods which systematically crawl social media platforms for posts containing a fact-checking label may be a more comprehensive and scalable approach to data collection [67]. This is again, limited to those claims identified by official fact-checkers, but the approach would allow multiple iterations of a claim to be recorded, making samples more representative of the health misinformation which exists in the wild. Similarly, rather than relying on literature, surveys could prove an effective way of discerning the health topics about which women are most concerned, and act as a more user-centred basis for topically focusing the sample.
7.2 Study 2

The sample of participants recruited for Study 2 was small, and not representative of the general population. Participants were initially recruited from the researcher’s social network, and even those recruited from the general public were mainly concentrated in the age range 25-35, and were university educated. Future work could recruit a more diverse sample in terms of age, educational background, and geographic location, to explore how misinformation is experienced by women in different countries. Furthermore, while some participants freely disclosed their disability status, this data was not requested. A dedicated and intersectional analysis of how disabled women experience health misinformation may have been useful, as the results suggest that they may be particularly vulnerable to misinformation about the conditions they live with.

The sample of posts collected was subject to the selection criteria of being identified as misinformation by participants. Though this was intentional, future work could pay dedicated attention to misinformation which goes undetected by users, as this may be tacit, and more likely to do harm. Moreover, the lack of direct access to posts made it difficult to verify the descriptions provided by participants, and the correctness of their source attributions. With ethical approval, future work could allow the submission of screenshots, to provide additional context, and to remove the burden of describing the post from the participant. Lastly, the diary procedure was relatively involved and required participants to fill in a questionnaire with free-entry text boxes. Though this facilitated rich and structured entries, this may have been at the expense of immediacy, as participants often waited until later to fill out the questionnaire. Experience sampling or more novel approaches to diary capture such as selfie-style videos [64] may facilitate more immediate data collection without sacrificing detail.

8. CONCLUSION

In summary, this dissertation reports two complementary studies. The first study involved a content analysis of social media posts related to women’s health marked as misinformation by official fact-checking organisations, and the second employed a diary and interview study to explore the situated sensemaking process underpinning women’s interactions with health misinformation on social media. Most importantly, the findings from both studies show that gendered dynamics are indeed at play in both the topical focus and framing of health misinformation on social media. Furthermore, affordances such as length constraints and content moderation were found to influence the contexts in which misinformation was encountered, and participants’ reactions to the material. No one feature or affordance was found to unilaterally help or hinder participants in making sense of misinformation, and outcomes depended on the type of sensemaking gaps experienced by participants, and on additional situational factors. Lastly, Study 2 draws attention to user-platform power imbalances, and how this is reflected in user wariness and distrust towards opaque algorithmic mediation. Therefore, as well as presenting a novel, multi-platform analysis of health misinformation which centres the perspectives of women, this dissertation poses some intriguing areas of future investigation where gender has traditionally been neglected, such as AI explainability and decentralised, community-led fact checking.

Acknowledgments

I would like to thank my supervisors Aneesha Singh and Dilisha Patel for their support, guidance and inspiration throughout all stages of the project, and their willingness to provide feedback. I would also like to extend my thanks to those who participated in Study 2, and took the time to record their experiences in detail. Lastly, I would like to thank my partner, friends, family, and my dog Loki for their moral support throughout the writing of the dissertation.

References

Appendix A: Codebook

1  CORE VARIABLES
For each fact-check article returned by the API, collect the following raw variables from the response if available.

1. **Claim Date:** The date on which the post was created
2. **Review Date:** The date on which the fact-check article was published
3. **Verdict:** The one or two-word rating assigned to the claim by the fact-checking organisation, e.g. ‘False’ or ‘Half-true’
4. **Article URL:** Link to the fact-checking article
5. **Name of fact-checking organisation:** E.g. Snopes, FullFact, etc.

2  CLAIMS
What is the key topic or claim of the misinformation?

1. **Reproductive health:** Misinformation surrounding contraception and abortion, menstruation, pregnancy, breastfeeding or any other topic relating to reproductive health.
2. **COVID-19 and vaccination:** Misinformation focusing on the COVID-19 virus or vaccine, and its effects on pregnancy, fertility, or breastfeeding.
3. **Cancer:** Bogus cancer treatments or cures, and misinformation about the causes of cancer.
4. **Dieting and weight loss:** Bogus weight loss products or remedies, or misinformation about nutrition and weight management.
5. **Actions of public health bodies:** Misinformation surrounding the decisions, actions, or statements of an official public body both recently and in the past (e.g. the FDA, WHO, NHS etc.)

3  TYPES
Based on the longer verdict, what type of misinformation is this?

1. **Misleading content:** Information which is factually inaccurate or incomplete, but has some grounding in existing facts and is not entirely fabricated. It may be partially true or reconfigured in some way.
2. **Missing context:** Information where the original context has been omitted or changed in some way, i.e., the content was intended as satire or information, or the post involves incorrect attributions to individuals or organisations.
3. **Misrepresentation of Data:** Flawed scientific reasoning, and misrepresentation of statistical data from reports, studies or otherwise. This includes mistaking a correlation for a causation, cherry-picking statistics, and other numerical fallacies.
4. **Fabricated content:** Information which is completely false or made up, with no factual grounding.
4 SOURCES AND CLAIMANTS

From which platform or website did the main post linked by the fact-checker originate?

1. Facebook
2. Instagram
3. Twitter
4. TikTok
5. YouTube
6. WhatsApp
7. A mainstream news source (provide name)
8. An alternative news source (provide name)
9. Other (provide name)

Judging by the posting account’s page, bio, or recent post history, select the best matching category or sub-category from the following:

1. Member of the public
2. Public figure or celebrity (specify whether mainstream or alternative/online)
3. Organisation or business
4. Other (specify details)

5 MEDIA FORMATS

Select the main media format through which the misinformative claim is expressed

1. Text: Information is disseminated primarily or exclusively through text. Include posts which contain images which are irrelevant to the content or not used to convey any informational or emotive content.
2. Image: Indicate whether the post is a screenshot of another social media post, article, or if it is a meme.
3. Video: Indicate whether the video is a TV broadcast or interview, an informal reel, or other
4. Digital article: Includes news articles, informal blogs, and think pieces

6 INTERVENTIONS

If possible to discern, what is the current status of the post?

1. Deleted or account suspended
2. Fact-checked by the platform
3. Still circulating without a disclaimer
Appendix B: Recruitment Flyer

UCLIC - University College London

Volunteers Needed For Research Investigating Women’s Experiences With Health Misinformation On Social Media

Are you a regular user of social media? Have you recently come across health misinformation online?

If so, you might be eligible to take part in valuable MSc. research focusing on how women experience and interpret health misinformation on social media

You may qualify if you:

- Identify as a woman
- Are aged 18+
- Speak English fluently
- Regularly use social media
- Feel that you are regularly exposed to health misinformation online

Participation Involves:

Making a log of 3-5 instances of coming across health-related misinformation on social media and participating in a follow-up interview lasting approximately 30-45 minutes.

Participation will be compensated with gift vouchers.

What are the benefits of taking part?

You will have an opportunity to contribute to a growing understanding of how health misinformation is interpreted by social media users and how it affects women. This may inform future design interventions to tackle misinformation.

FOR MORE INFORMATION

Please contact [email protected] with the subject ‘Women’s Health and Misinformation Study’
1 DEMOGRAPHICS

(1) What is your age group?
   - 18–24
   - 25–34
   - 35–44
   - 45–54
   - 55–64
   - Prefer not to answer

(2) How would you describe your gender?
   - Male
   - Female
   - Non-binary
   - Other, please specify ___________________________________________
   - Prefer not to answer

(3) What is your ethnic group?
   - White
   - Asian or Asian British
   - Black British, Caribbean, or African
   - Arab or North African
   - Mixed or multiple ethnic groups
   - Other ethnic group, please specify ___________________________________________
   - Prefer not to answer

(4) In what sector do you work?
   - Healthcare
   - Information and Technology
   - Education
   - Retail and Hospitality
   - Business and Finance
   - Other, please specify ___________________________________________

(5) What is your highest level of education?
   - Secondary School
   - College/Sixth Form
   - University (undergraduate)
   - University (postgraduate)

(6) If you went to, or currently are attending university, which subject or discipline did you study and at what level? (E.g., BSc. Mathematics or MA English Literature)

____________________________________________________________________________________________
2 SOCIAL MEDIA HABITS

(1) Roughly how long do you spend on social media per day?
   ○ Less than 30 minutes
   ○ Between 30 minutes and an hour
   ○ Between 1 and 2 hours
   ○ Between 2 and 4 hours
   ○ Between 4 and 6 hours
   ○ More than 6 hours

(2) Which social media platforms do you use regularly? Please select all that apply.
   □ Facebook
   □ WhatsApp
   □ Twitter
   □ Instagram
   □ Snapchat
   □ Reddit
   □ TikTok
   □ Other, please specify ____________________________________________

3 DIGITAL AND INFORMATION LITERACY

For each of the following statements, please select the extent to which you agree.

<table>
<thead>
<tr>
<th>Statement</th>
<th>Strongly disagree</th>
<th>Somewhat disagree</th>
<th>Neither agree nor disagree</th>
<th>Somewhat agree</th>
<th>Strongly agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>I am a competent user of technology</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>I trust the information provided by official or mainstream sources</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>I find it easy to identify false or misleading information on social media</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>I always check that information is correct or accurate before sharing it on social media</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>I often feel overwhelmed or burnt out by the content I encounter on social media</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>I feel like I encounter health-related misinformation on social media often</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
</tbody>
</table>
Appendix D: Pre-study Interview Topic Guide

Questions to be asked after the participant is given a brief overview of the study’s aims and procedures.

1. How often do you use the Internet to seek out health information?
2. Are there any particular health topics or issues that you find particularly important or receive information about often? If so, what are they, and why?
3. To what extent do you trust the health information you encounter on the Internet, and why?
4. What comes to mind when you hear the word ‘misinformation’?
5. Can you think of any examples of health misinformation that you’ve encountered online recently?
   a. What makes this an example of misinformation?
   b. What impact did seeing this have on you?
   c. What do you think was the intent of the person who made the claim?
   d. Do you think intent matters?
6. How confident do you feel in your ability to identify misinformation online?
7. What social media platforms do you use the most often?
8. How well do you think those social media platforms manage misinformation, and why?

Appendix E: Diary Study Briefing Sheet

[PARTICIPANT ID AND LINKS TO SURVEYS HERE]

1. WHEN SHOULD I RECORD AN ENTRY?

Record an entry whenever you come across health-related content online/on social media that matches any of the following criteria:

- You feel that it contains misinformation - i.e., information that is false or exaggerated
- You find the content provocative or surprising.

Health topics may include (but are not limited to) reproductive and sexual health, mental and psychological health, beauty and body image, dieting, fitness, COVID-19, vaccinations (for COVID-19 or otherwise).

2. HOW SHOULD I RECORD AN ENTRY?

Following the URL at the top of the page will take you to the diary entry questionnaire. Simply click the link and fill it in whenever you want to record a new entry, being sure to enter your participant ID correctly for every entry. This is important, as identifying your entries will be very difficult without it.

You should aim to record an entry as soon as possible after interacting with the post. If you’re not able to fill in the questionnaire immediately, bookmark or screenshot the post when you come across it and take brief notes on your phone. Later, when you have the chance, use the screenshot/notes to help you fill in the diary questionnaire which you can access using the link at the top of this document.

While you are not expected to submit any screenshots, please keep a record of the posts you submit (either as bookmarks or screenshots on your phone). You will be interviewed on your entries later, so keeping a record for yourself may help you answer the interview questions.
3 HOW OFTEN SHOULD I RECORD ENTRIES?
We’re looking for at least 5 entries (but you are free to provide more) to be submitted over the period of 2 weeks.

4 HOW MUCH DETAIL SHOULD I PROVIDE IN MY ENTRIES?
Many of the diary questions are multiple choice, but for the free entry questions, we are looking for a few sentences/a short paragraph. You don’t need to write an essay, but please try to cover all the points in the prompts if applicable to you in reasonable detail- i.e., not one-word answers! Remember, the three best quality responses will win an extra £5 in vouchers, so a high-quality diary means you have a better chance of winning!

5 WHAT SHOULD I DO IF I HAVE ANY MORE QUESTIONS?
You are free to contact the researcher(s) at any point during the study, via the email addresses provided on the study information sheet.
Appendix F: Diary Study Questionnaire (Adapted from Qualtrics)

(1) On which social media platform or website did you see the content? If it was linked or shared from another website, please also specify the site of origin.
______________________________________________________________________________________________

(2) What were you doing at the time you came across the content?
- Looking for specific information
- Communicating with a friend or family member
- Nothing in particular/just passing time
- Other, please specify ___________________________________________

(3) To the best of your knowledge, what type of account posted the content?
- A public figure, such as a politician or celebrity
- An official public body or organisation (e.g. WHO)
- A business or media company
- A family member, friend, or acquaintance
- A member of the public that I don’t know or follow
- I don’t know
- Other, please specify ___________________________________________

(4) Roughly how long did you spend engaging with the content before you moved on?
- Less than 30 seconds
- Between 30 seconds and 1 minute
- Between 2 and 4 minutes
- Between 5 and 10 minutes
- Over 10 minutes

(5) In your own words, provide a brief description of the content, and why you found it thought provoking, or possibly misinformative. Please be sure to specify its media format (e.g., whether it is a video, article, advertisement, or text post).

______________________________________________________________________________________________
(6) If applicable, what questions, concerns, or uncertainties did the content raise?

(7) Did you investigate the content further? Why/why not?

(8) If you investigated the content further, please describe how you did so step-by-step, and specify any additional sources of information you used and how/whether they helped.

(9) Were there any specific features of the website, or social media platform you were on that helped you make sense of the content? If so, how?
Appendix F: Diary Study Questionnaire (Adapted from Qualtrics)

(10) Did you take any further steps such as sharing the content with someone else or reporting it? Why/Why not?

(11) Please add any further comments you feel are relevant to your entry.
Appendix G: Follow-up Interview Topic Guide

1 GENERAL REFLECTIONS

(1) How did you find participating in the diary study overall?
(2) Were there any parts of the study you found difficult?
(3) If you could do the study again, would you do anything differently?
(4) Which entry was the most memorable to you, and why?

2 DIARY-SPECIFIC PROBES

Questions specific to participants’ diaries are asked here. Below are some generic, commonly asked questions, but the interviews were specialised to each participant’s diaries so varied significantly.

Situational Context

(1) If content was from a followed account/friend: do you encounter content from this individual or account often?
   (a) If so, what sort of content do they usually post?
   (b) How did this content fit into what they usually post?
(2) If post was in a group: Why are you a member of this group, and what do you use it for?
(3) If else: why do you think this post appeared on your feed?

Gaps

(1) How important was it that you find an answer to this question/uncertainty, and why?
(2) How easy or difficult did you expect answering this question to be?
(3) Was there anything standing in the way of you answering this question?

Bridges

(1) What motivated you to use this strategy?
(2) To what extent were you expecting this strategy to help you?
(3) How easy or difficult was it to use this strategy?

Outcomes

(1) Did [bridging strategy] work as expected? Did it provide the expected clarity/answer to your question?
   (a) If not, why do you think that was the case?
   (b) What do you think would have helped in retrospect?
(2) Have you thought about any of the posts you encountered in the time since logging your entry?
   (a) If so, under what circumstances?
   (b) Have your feelings towards the post changed or evolved since first encountering it?
Appendix H: Information Sheet and Consent Form

**Title of Study**: Investigating how women make sense of health misinformation on social media

**Department**: UCLIC

**Researcher(s):**

**Principal Researcher**: Aneesha Singh ([aneesha.singh@ucl.ac.uk](mailto:aneesha.singh@ucl.ac.uk))

This study has been approved by the UCLIC Research Ethics Committee, Project number: UCLIC_1920_007_Staff_Singh

1. **Invitation Paragraph**

   You are being invited to take part in my MSc. Research project. Before you decided it is important for you to understand why this study is being done and what participation will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether you wish to take part. Thank you for reading this.

2. **What is the project's purpose?**

   The aim of this project is to better understand how women and people assigned female at birth are affected by health misinformation on social media, and how they make sense of false or misleading posts. This includes their strategies for identifying misinformation, and the emotional and social factors which affect this process, with a view to supporting the design of technologies which support the identification and management of misinformation by users.

3. **Why have I been chosen?**

   You have been invited to participate because you:

   - Identify as a woman or are assigned female at birth (AFAB)
   - Are over the age of 18
   - Are a regular user of social media and other social digital platforms
   - Can communicate effectively in English and do not consider yourself to be a vulnerable adult.
   - Can give informed consent.

   I am aiming to recruit between 10-15 participants in total to participate in this study.

4. **Do I have to take part?**

   It is up to you to decide whether or not to take part. If you do decide to take part you will be given this information sheet to keep and be asked to sign a consent form. You can withdraw at any time without giving a reason. If you decide to withdraw you will be asked what you wish to happen to the data you have provided up that point.

5. **What will happen to me if I take part?**

   You will be invited to participate in an initial interview, in which the purpose and methods of the study will be explained, and you will be asked a few questions about your encounters with misinformation online. This session should take no more than 20 minutes and will give you an opportunity to ask any questions about the study. If you agree to continue with the study after this, you will be sent a brief survey which will collect
basic demographical information, details on your social media habits and your prior experience of health misinformation on social media. This will take approximately 5 minutes to complete.

Following submission of your survey, you will be asked to record at least 3 accounts of encountering health misinformation on social media using a structured questionnaire, the URL of which will be provided to you. You are expected to go about your daily social media usage habits and record an entry whenever you come across a post you feel broadly contains health misinformation, including posts that have been officially or unofficially fact checked by organisations or other users.

After you have submitted your entries, you will be invited to a final debriefing interview in which your experience of the diary study will be discussed, and key moments from your entries will be explored in more detail. This should take between 30-45 minutes. You will be compensated £15 for participation in the study, and an additional £5 will be available for each of the five most detailed diary entries that are submitted. If you withdraw, you will be compensated partially, depending on how much of the study you complete.

6. Will I be recorded and how will the recorded media be used?

With your permission, the interviews will be audio recorded. Transcriptions of the audio recording and your online diary and survey data will be used only for analysis and for illustration in conference presentations and lectures. The audio recordings will be deleted once they have been transcribed and any identifying information will be removed during transcription.

7. What are the possible disadvantages and risks of taking part?

No disadvantages or risks of taking part have been identified. In the unlikely event that participating causes you any distress, you are free to withdraw, to discuss concerns with the researcher or the Principal Investigator.

8. What are the possible benefits of taking part?

While there are no immediate benefits to you from taking part, we hope that you will find the study interesting and that it will help you to reflect on how you identify and interact with health misinformation on social media.

We aim to share our findings with developers of future apps in this space so that they may help to inform future interventions aimed at curtailing health misinformation on social media.

9. What if something goes wrong?

If you have any concerns with the conduct of this study, please raise them in the first instance with Dr Aneesha Singh (aneesha.singh@ucl.ac.uk). If your concerns are not addressed to your satisfaction then you may contact the Chair of the UCL Research Ethics Committee – ethics@ucl.ac.uk

10. Will my taking part in this project be kept confidential?

All the information that we collect about you during the course of the research will be kept confidential, subject to legal constraints and professional guidelines. You will not be identifiable in any ensuing reports or publications.

11. What will happen to the results of the research project?

This study is for my MSc. project, and the findings will be reported in my dissertation. If you would like to receive a copy of that, let me know and I will send it to you in August / September. Depending on the findings, my supervisor and I may also publish the results in a journal or conference paper. Anonymised data will be stored securely for five years and may be reviewed in subsequent studies that have a related focus.
12. Local Data Protection Privacy Notice

The controller for this project will be University College London (UCL). The UCL Data Protection Officer provides oversight of UCL activities involving the processing of personal data, and can be contacted at data-protection@ucl.ac.uk

The only personal information retained will be a copy of your informed consent and your chosen contact details if you wish to be informed of the outcome of this study. These will be held securely and separately from the anonymised data that you provide for the study.

Further information on how UCL uses participant information can be found at https://www.ucl.ac.uk/legal-services/privacy/ucl-general-research-participant-privacy-notice

The lawful basis that would be used to process your personal data will be performance of a task in the public interest.

Your personal data will be processed so long as it is required for the research project. If we are able to anonymise or pseudonymise the personal data you provide we will undertake this, and will endeavour to minimise the processing of personal data wherever possible.

If you are concerned about how your personal data is being processed, or if you would like to contact us about your rights, please contact UCL in the first instance at data-protection@ucl.ac.uk.

13. Contact for further information

Contact details for me and my supervisor are provided at the top of this sheet; feel free to contact either of us if you have queries or concerns.

Thank you for reading this information sheet and for considering taking part in this study.
CONSENT FORM FOR ADULTS ENCOUNTERING HEALTH MISINFORMATION IN RESEARCH STUDIES

Please complete this form after reading the Information Sheet or listening to an explanation of the study.

Thank you for considering taking part in this research. If you have any questions arising from the Information Sheet or explanation already given to you, please ask the researcher before you decide whether to join in. You will be given a copy of this Consent Form to keep and refer to.

I confirm that I understand that by ticking each box below I am consenting to this element of the study. I understand that unticked boxes means that I DO NOT consent to that part of the study. I understand that by not giving consent for certain elements, I may be deemed ineligible for the study.

<table>
<thead>
<tr>
<th></th>
<th>Tick</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>I confirm that I have read and understood the Information Sheet for the above study. I have had an opportunity to consider the information and what will be expected of me and to ask questions which have been answered to my satisfaction. I agree to take part.</td>
</tr>
<tr>
<td>2.</td>
<td>I understand that data will be anonymised and that it will not be possible to link my personal data (consent, contact details) with the study data. I understand that according to data protection legislation, ‘public task’ will be the lawful basis for processing.</td>
</tr>
<tr>
<td>3.</td>
<td>I understand that all personal information will remain confidential and that all efforts will be made to ensure I cannot be identified. Data gathered in this study will be stored pseudonymously and securely. It will not be possible to identify me in any publications.</td>
</tr>
<tr>
<td>4.</td>
<td>I understand that my information may be subject to review by responsible individuals from the University for monitoring and audit purposes.</td>
</tr>
<tr>
<td>5.</td>
<td>I understand the direct/indirect benefits of participating.</td>
</tr>
<tr>
<td>6.</td>
<td>I understand that I will not benefit financially from this study or from any possible outcome it may result in in the future.</td>
</tr>
<tr>
<td>7.</td>
<td>I understand that I will be compensated for the portion of time spent in the study if I choose to withdraw.</td>
</tr>
<tr>
<td>8.</td>
<td>I agree that my anonymised research data may be used by others for future research. [No one will be able to identify you when this data is shared.]</td>
</tr>
<tr>
<td>9.</td>
<td>I understand that the information I have submitted will be published as a report.</td>
</tr>
<tr>
<td>10.</td>
<td>I consent to my interview being audio/video recorded and understand that the recordings will be destroyed/ pseudonymised following transcription.</td>
</tr>
<tr>
<td>11.</td>
<td>I confirm that I understand the inclusion and exclusion criteria as detailed in the Information Sheet and explained to me by the researcher and that I fall under the inclusion criteria and I do not fall under the exclusion criteria.</td>
</tr>
<tr>
<td>12.</td>
<td>I am aware of who I should contact if I wish to lodge a complaint.</td>
</tr>
</tbody>
</table>

_________________________ ________________ ___________________
Name of participant Date Signature